Exact recursive evaluation of 3*j*- and 6*j*-coefficients for quantum-mechanical coupling of angular momenta*

Klaus Schulten† and Roy G. Gordon

Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138
(Received 21 April 1975)

Algorithms are developed for the exact evaluation of the 3*j*-coefficients of Wigner and the 6*j*-coefficients of Racah. These coefficients arise in the quantum theory of coupling of angular momenta. The method is based on the exact solution of recursion relations in a particular order designed to guarantee numerical stability even for large quantum numbers. The algorithm is more efficient and accurate than those based on explicit summations, particularly in the commonly arising case in which a whole set of related coefficients is needed.

I. INTRODUCTION

Common algorithms for the evaluation of 3*j*- and 6*j*-coefficients are based on the explicit expressions of Wigner[1] and Racah.[2] Calculations involving the quantum mechanical coupling of angular momenta often require the evaluation of whole strings of coupling coefficients of the kind:

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \quad \text{for all allowed } j_1, \tag{1}$$

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & -m_1-m_2 \end{pmatrix} \quad \text{for all allowed } m_2, \tag{2}$$

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix} \quad \text{for all allowed } j_1. \tag{3}$$

Numerical examples of these sets of coupling coefficients are given in Figs. 1–3.

The existing algorithms, however, evaluate coupling coefficients separately and do not make use of relationships between the values of neighboring 3*j*- and 6*j*-coefficients. The algorithms, furthermore, are inapplicable for large angular momentum values ($\sim 100\hbar$) which, for example, occur frequently in problems of molecular dynamics.

We have now numerically tested an algorithm for the evaluation of 3*j*- and 6*j*-coefficients based on recursion equations relating the coefficients in the strings (1), (2), or (3). This algorithm simultaneously generates all coupling coefficients within these strings without more numerical effort than is needed to evaluate a single coupling coefficient. Further, this algorithm is numerically applicable for large angular momentum quantum numbers.

In the following, we will present the derivations of the recursion equations which relate the coupling coefficients in (1), (2), or (3). In Sec. II we derive these recursion relations algebraically from certain sum rules satisfied by these coefficients. While this derivation is the shortest available, it is somewhat remote from the definitions of the coefficients. Thus in the Appendix, we supply an alternate derivation starting directly from the basic definitions of angular momentum coupling. In Sec. III we then derive the algorithm for generating the strings of 3*j*- and 6*j*-coefficients (1), (2), and (3). In Sec. IV we demonstrate numerically the accuracy and efficiency of the algorithm. Computer programs for the recursive evaluation of 3*j*- and 6*j*-coefficients will be made available.[3]

Beside being most advantageous for numerical evaluations, the recursion equations serve to make the functional properties of the angular momentum coupling coefficients more transparent. In a second article following this one,[4] it is shown that the recursion equations for 3*j*- and 6*j*-coefficients can be solved using a discrete analog of the uniform WKB approximation to yield simple analytic approximate expressions for individual coupling coefficients, which are quite accurate even for moderate quantum numbers.

II. RECURSION RELATIONSHIPS FOR 3*j*- AND 6*j*-COEFFICIENTS

The recursion relationships which connect the angular momentum coupling coefficients in (1), (2), and (3) had been previously reported. Condon and Shortley[5] derived the recursion relationships for the 3*j*-coefficients in (1), and Rose[6] presented the recursion relationship for the 3*j*-coefficients in (2). In both instances the recursion relationships were obtained from the interpretation of the strings of 3*j*-coefficients in (1) and (2) as the eigenvectors of certain angular momentum operators. Condon and Shortley, and subsequently Rose, suggested that these recursion equations might help evaluate the 3*j*-coefficients. The recursion equation for the 6*j*-coefficients in (3) have been given by Yutsis *et al.*[7] In an appendix following this paper, we show that this recursion equation, too, originates from an eigenvalue problem. Instead of now just quoting the recursion equations of Condon and Shortley, Rose and Yutsis *et al.*, we present a unified derivation for these three recursion equations. This derivation starts off from three basic sum rules which hold for 3*j*- and 6*j*-coefficients.

Let us first consider the 3*j*-coefficients in (1). For the 3*j*-coefficients there is an identity[8]

$$(-1)^{j_2+m_2+j_3-m_3+l_1+m_1'} \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \begin{pmatrix} l_1 & l_2 & j_3 \\ m_1' & m_2' & -m_3 \end{pmatrix}$$

$$= \sum_{l_3} (2l_3+1) \begin{pmatrix} j_1 & l_2 & l_3 \\ m_1 & m_2' & m_1'+m_2 \end{pmatrix} \begin{pmatrix} l_1 & j_2 & l_3 \\ -m_1' & -m_2 & m_1'+m_2 \end{pmatrix}$$

$$\times \begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix}. \tag{4}$$

For $l_1=\frac{1}{2}$, $l_2=j_3+\alpha$, and $m_1'=\beta$ $(\alpha,\beta=\pm\frac{1}{2})$ this identity reduces to the three term recursion relationship

$$(-1)^{j_2+m_2+j_3-m_3+1/2+\beta} \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & j_3+\alpha & j_3 \\ \beta & m_3-\beta & -m_3 \end{pmatrix}$$

$$= \sum_{l_3=j_2-1/2}^{j_2+1/2} (2l_3+1) \begin{pmatrix} j_1 & j_3+\alpha & l_3 \\ m_1 & m_3-\beta & m_2+\beta \end{pmatrix}$$

$$\times \begin{pmatrix} \frac{1}{2} & j_2 & l_3 \\ -\beta & -m_2 & m_2+\beta \end{pmatrix} \begin{Bmatrix} j_1 & j_2 & j_3 \\ \frac{1}{2} & j_3+\alpha & l_3 \end{Bmatrix} \tag{4'}$$

which connects the $3j$-coefficients

$$\begin{pmatrix} j_1 & j_2-\frac{1}{2} & j_3+\alpha \\ m_1 & m_2+\beta & m_3-\beta \end{pmatrix}, \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}$$

and

$$\begin{pmatrix} j_1 & j_2+\frac{1}{2} & j_3+\alpha \\ m_1 & m_2+\beta & m_3-\beta \end{pmatrix}.$$

The factors multiplying these $3j$-coefficients in Eq. (4′) are $3j$- and $6j$-coefficients containing a quantum number $\frac{1}{2}$ for which closed expressions exist. Equation (4′) with $\alpha=-\frac{1}{2}$ and $\beta=-\frac{1}{2}$ is identical with a recursion relationship previously derived by Louck[9] starting from the Clebsch–Gordan series.

Recursion relationships (5) properly combined give a recursion relationship which only connects $3j$-coefficients belonging to (1). To be specific, the recursion relationships to be combined are

$$(-1)^{2m_3} (2j_2+1)\,[j_3-m_3+1]^{1/2} \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}$$

$$= [(j_1+j_2-j_3)(j_1-j_2+j_3+1)(j_2-m_2)]^{1/2}$$

$$\times \begin{pmatrix} j_1 & j_2-\frac{1}{2} & j_3+\frac{1}{2} \\ m_1 & m_2+\frac{1}{2} & m_3-\frac{1}{2} \end{pmatrix}$$

$$-[(j_1+j_2+j_3+2)(-j_1+j_2+j_3+1)(j_2+m_2+1)]^{1/2}$$

$$\times \begin{pmatrix} j_1 & j_2+\frac{1}{2} & j_3+\frac{1}{2} \\ m_1 & m_2+\frac{1}{2} & m_3-\frac{1}{2} \end{pmatrix}, \tag{5a}$$

$$(-1)^{2m_3} 2j_2[j_3-m_3+1]^{1/2} \begin{pmatrix} j_1 & j_2-\frac{1}{2} & j_3+\frac{1}{2} \\ m_1 & m_2+\frac{1}{2} & m_3-\frac{1}{2} \end{pmatrix}$$

$$= -[(j_1+j_2+j_3+1)(-j_1+j_2+j_3)(j_2+m_2)]^{1/2}$$

$$\times \begin{pmatrix} j_1 & j_2-1 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}$$

$$+[(j_1-j_2+j_3+1)(j_1+j_2-j_3)(j_2-m_2)]^{1/2}$$

$$\times \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}, \tag{5b}$$

and

$$(-1)^{2m_3} (2j_2+2)[j_3-m_3+1]^{1/2} \begin{pmatrix} j_1 & j_2+\frac{1}{2} & j_3+\frac{1}{2} \\ m_1 & m_2+\frac{1}{2} & m_3-\frac{1}{2} \end{pmatrix}$$

$$= -[(j_1+j_2+j_3+2)(-j_1+j_2+j_3+1)(j_2+m_2+1)]^{1/2}$$

$$\times \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}$$

$$+[(j_1-j_2+j_3)(j_1+j_2-j_3+1)(j_2-m_2+1)]^{1/2}$$

$$\times \begin{pmatrix} j_1 & j_2+1 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}. \tag{5c}$$

Inserting (5b) and (5c) into (5a) gives a recursion relationship for $3j$-coefficients which may be written

$$j_1 A(j_1+1) \begin{pmatrix} j_1+1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} + B(j_1) \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}$$

$$+ (j_1+1) A(j_1) \begin{pmatrix} j_1-1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = 0, \tag{6a}$$

where

$$A(j_1) = [j_1^2-(j_2-j_3)^2]^{1/2}[(j_2+j_3+1)^2-j_1^2]^{1/2}[j_1^2-m_1^2]^{1/2}, \tag{6b}$$

$$B(j_1) = -(2j_1+1)[j_2(j_2+1)m_1 - j_3(j_3+1)m_1 - j_1(j_1+1)(m_3-m_2)]. \tag{6c}$$

Recursion equation (6), it will be shown below, together with the normalization condition

$$\sum_{j_1} (2j_1+1) \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}^2 = 1, \tag{7}$$

is sufficient to determine except for an overall phase factor the values of the $3j$-coefficients in (1).

There exists yet another recursion equation for $3j$-coefficients, which relates $3j$-coefficients with different magnetic quantum numbers, and which allows the evaluation of the elements in (2). This recursion equation is derived in much the same manner as Eq. (6). Hence, we may only outline this derivation. It had already been pointed out by Edmonds[10] that the identity

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix}$$

$$= \sum_m (-1)^{\psi_m} \begin{pmatrix} j_1 & l_2 & l_3 \\ m_1 & m & -m_1-m \end{pmatrix} \begin{pmatrix} l_1 & j_2 & l_3 \\ m_3-m & m_2 & m_1+m \end{pmatrix}$$

$$\times \begin{pmatrix} l_1 & l_2 & j_3 \\ m-m_3 & -m & m_3 \end{pmatrix}, \tag{8}$$

$$\psi_m = l_1+l_2+l_3+m_1-m_3-m,$$

provides a suitable starting point for the derivation of

recursion relationships for $3j$-coefficients. Setting $l_1 = \frac{1}{2}$, $l_2 = j_3 + \beta$, and $l_3 = j_2 + \alpha$ $(\alpha, \beta = \pm\frac{1}{2})$ gives the three term recursion relationship

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \begin{Bmatrix} j_1 & j_2 & j_3 \\ \frac{1}{2} & j_3+\beta & j_2+\alpha \end{Bmatrix}$$

$$= \sum_{m=m_3-1/2}^{m_3+1/2} (-1)^{\phi_m} \begin{pmatrix} j_1 & j_3+\beta & j_2+\alpha \\ m_1 & m & -m_1-m \end{pmatrix}$$

$$\times \begin{pmatrix} \frac{1}{2} & j_2 & j_2+\alpha \\ m_3-m & m_2 & m_1+m \end{pmatrix} \begin{pmatrix} \frac{1}{2} & j_3+\beta & j_3 \\ m-m_3 & -m & m_3 \end{pmatrix} \tag{8'}$$

which connects the $3j$-coefficients

$$\begin{pmatrix} j_1 & j_2+\alpha & j_3+\beta \\ m_1 & m_2+\frac{1}{2} & m_3-\frac{1}{2} \end{pmatrix}, \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \text{ and}$$

$$\begin{pmatrix} j_1 & j_2+\alpha & j_3+\beta \\ m_1 & m_2-\frac{1}{2} & m_3+\frac{1}{2} \end{pmatrix}.$$

The factors multiplying these $3j$-coefficients are again $3j$- and $6j$-coefficients containing a quantum number $\frac{1}{2}$ for which closed expressions exist. From (8') can then be obtained by a proper combination of three recursion relationships the following equation which relates the $3j$-coefficients belonging to (2)

$$C(m_2+1)\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2+1 & m_3-1 \end{pmatrix} + D(m_2)\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}$$

$$+ C(m_2)\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2-1 & m_3+1 \end{pmatrix} = 0 \tag{9a}$$

where

$$C(m_2) = [(j_2 - m_2 + 1)(j_2 + m_2)(j_3 + m_3 + 1)(j_3 - m_3)]^{1/2}, \tag{9b}$$

$$D(m_2) = j_2(j_2+1) + j_3(j_3+1) - j_1(j_1+1) + 2m_2m_3. \tag{9c}$$

It will be shown that Eq. (9) together with the normalization condition

$$\sum_{m_2} (2j_1+1)\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & -m_1-m_2 \end{pmatrix}^2 = 1 \tag{10}$$

is sufficient to determine except for an overall phase factor the values of the $3j$-coefficients in (2).

The recursion equation which selectively connects the $6j$-coefficients belonging to the set (3) is derived in a manner strikingly similar to the recursion equations (6) and (9) above. Now the Biedenharn–Elliot identity[11] serves as the starting point:

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix} \begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1' & l_2' & l_3' \end{Bmatrix}$$

$$= \sum_{\lambda} (-1)^{\phi_\lambda} \begin{Bmatrix} j_1 & l_2' & l_3' \\ \lambda & l_3 & l_2 \end{Bmatrix} \begin{Bmatrix} l_1' & j_2 & l_3' \\ l_3 & \lambda & l_1 \end{Bmatrix} \begin{Bmatrix} l_1' & l_2' & j_3 \\ l_2 & l_1 & \lambda \end{Bmatrix}, \tag{11}$$

$$\phi_\lambda = j_1 + j_2 + j_3 + l_1 + l_2 + l_3 + l_1' + l_2' + l_3' + \lambda.$$

If one sets $l_1' = \frac{1}{2}$, $l_2' = j_3 + \beta$, and $l_3' = j_2 + \alpha$ $(\alpha, \beta = \pm\frac{1}{2})$, the sum over λ reduces to two terms with $\lambda = l_1 \pm \frac{1}{2}$. One arrives then at the recursion relationship

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix} \begin{Bmatrix} j_1 & j_2 & j_3 \\ \frac{1}{2} & j_3+\beta & j_2+\alpha \end{Bmatrix}$$

$$= \sum_{\lambda=l_1-1/2}^{l_1+1/2} (-1)^{\phi_\lambda} \begin{Bmatrix} j_1 & j_2+\alpha & j_3+\beta \\ \lambda & l_2 & l_3 \end{Bmatrix} \begin{Bmatrix} \frac{1}{2} & j_2 & j_2+\alpha \\ l_3 & \lambda & l_1 \end{Bmatrix}$$

$$\times \begin{Bmatrix} \frac{1}{2} & j_3+\beta & j_3 \\ l_2 & l_1 & \lambda \end{Bmatrix} \tag{11'}$$

which connects the $6j$-coefficients

$$\begin{Bmatrix} j_1 & j_2+\alpha & j_3+\beta \\ l_1-\frac{1}{2} & l_2 & l_3 \end{Bmatrix}, \begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix} \text{ and}$$

$$\begin{Bmatrix} j_1 & j_2+\alpha & j_3+\beta \\ l_1+\frac{1}{2} & l_2 & l_3 \end{Bmatrix}.$$

The factors in this recursion relationship consist of $6j$-coefficients with a quantum number $\frac{1}{2}$ for which closed expressions exist. Proper combination of three recursion relationships (11') yields

$$j_1 E(j_1+1) \begin{Bmatrix} j_1+1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix} + F(j_1) \begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix}$$

$$+ (j_1+1) E(j_1) \begin{Bmatrix} j_1-1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix} = 0 \tag{12a}$$

where

$$E(j_1) = \{[j_1^2 - (j_2 - j_3)^2][(j_2 + j_3 + 1)^2 - j_1^2][j_1^2 - (l_2 - l_3)^2]$$
$$\times [(l_2 + l_3 + 1)^2 - j_1^2]\}^{1/2}, \tag{12b}$$

$$F(j_1) = (2j_1+1)\{j_1(j_1+1)[-j_1(j_1+1) + j_2(j_2+1) + j_3(j_3+1)]$$
$$+ l_2(l_2+1)[j_1(j_1+1) + j_2(j_2+1) - j_3(j_3+1)]$$
$$+ l_3(l_3+1)[j_1(j_1+1) - j_2(j_2+1) + j_3(j_3+1)]$$
$$- 2j_1(j_1+1)l_1(l_1+1)\}.$$

Recursion equation (12) together with the normalization condition

$$\sum_{j_1} (2j_1+1)(2l_1+1)\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix}^2 = 1 \tag{13'}$$

is sufficient to determine expect for an overall phase factor the $6j$-coefficients in (3).

Racah[11] had pointed out that his explicit formula is not the only pathway for an evaluation of $6j$-coefficients, but that instead the recursion equation (11') equally well furnishes an approach to the evaluation of $6j$-coefficients. Racah and Fano noted that the coefficients in these recursion equations consisting of $6j$-coefficients with quantum numbers $\frac{1}{2}$ are determined through the unitary property

$$\sum_{j_1} (2j_1+1)(2l_1+1)\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix}\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1' & l_2 & l_3 \end{Bmatrix} = \delta_{l_1 l_1'} \tag{13}$$

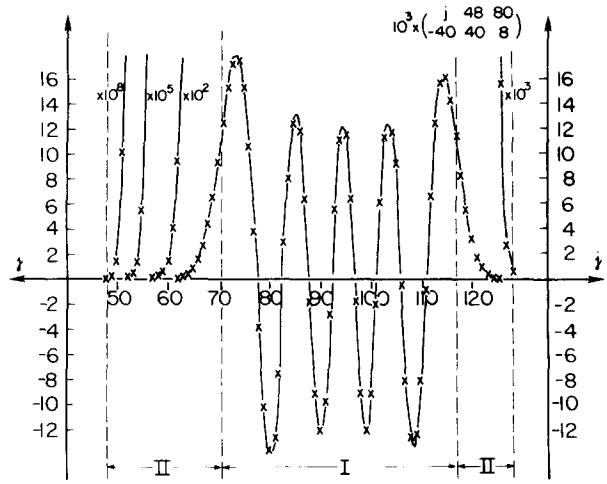


FIG. 1. Functional behavior of $3j$-coefficients $f(j_1)=\begin{pmatrix} j_1 & 48 & 80 \\ -40 & 40 & 8 \end{pmatrix}$ $(48 \le j_1 \le 128)$. The evaluation of $f(j_1)$ followed the recursion algorithm described in Sec. 3. The domain of $f(j_1)$ can be divided into a classical and two nonclassical regions (Ref. 4). In the classical region $f(j_1)$ oscillates with slowly varying amplitude; in the nonclassical regions $|f(j_1)|$ monotonically decays to zero.

together with the associative property

$$\sum_{j_1} (-1)^{j_1+l_1+l_1'} (2j_1+1) \begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix} \begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1' & l_3 & l_2 \end{Bmatrix} = \begin{Bmatrix} j_2 & l_3 & l_1 \\ j_3 & l_2 & l_1' \end{Bmatrix}. \tag{14}$$

Hence, the very interesting conclusion can be drawn that the identities (13), (14), and (11) completely determine the $6j$-coefficients save an overall phase factor.[11]

The similarity between the recursion equations (6) and (9) for $3j$-coefficients and the recursion equation (12) for $6j$-coefficients as well as the similarity between the corresponding derivations is explained through the existence of an asymptotic relationship between $3j$- and $6j$-coefficients[12]:

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = \lim_{R\to\infty} (-1)^{2l_1-2j_1} [2l_1+2R+1]^{1/2} \times \begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1+R & l_2+R & l_3+R \end{Bmatrix} \tag{15}$$

where $l_3-l_2=m_1$, $l_1-l_3=m_2$, and $l_2-l_1=m_3$. In fact, Eq. (6) follows from Eq. (12) by taking the asymptotic limit letting l_1, l_2, and l_3 go to infinity, whereas Eq. (9) follows from Eq. (12) by letting j_1, j_2, and l_3 go to infinity. We have chosen the series of $3j$- and $6j$-coefficients in Figs. 1, 2, and 3 to be related through the asymptotic relationship (15) as may be readily checked. The similarity of these diagrams is therefore an illustration for Eq. (15).

III. ALGORITHM FOR THE RECURSIVE EVALUATION OF $3j$- AND $6j$-COEFFICIENTS

The three-term recursion equations for $3j$- and $6j$-coefficients (6), (9), and (12) have been derived and it will now be shown how the Wigner and Racah coefficients can be determined from these recursion equations.

To describe the proposed recursive algorithm, we will first consider the evaluation of the string of $3j$-coefficients

$$f(j_1)=\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}, \quad j_{1\min} \le j_1 \le j_{1\max}. \tag{1'}$$

The range of j_1 is finite, the smallest and largest values being

$$j_{1\min}=\max\{|j_2-j_3|, \ |m_1|\} \text{ and } j_{1\max}=j_2+j_3.$$

Once proper starting values have been given, the re-

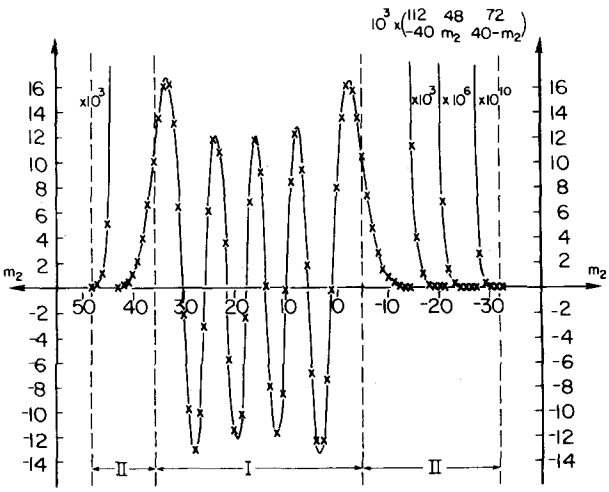


FIG. 2. Functional behavior of the $3j$-coefficients $g(m_2)$ $=\begin{pmatrix} 112 & 48 & 72 \\ -40 & m_2 & 40-m_2 \end{pmatrix}$ $(-32 \le m_2 \le 48)$. The evaluation of $g(m_2)$ followed the recursion algorithm described in Sec. 3. The domain of $g(m_2)$ can be divided into a classical region and two nonclassical regions (Ref. 4). In the classical region $g(m_2)$ oscillates with slowly varying amplitude; in the nonclassical regions $|g(m_2)|$ monotonically decays to zero.

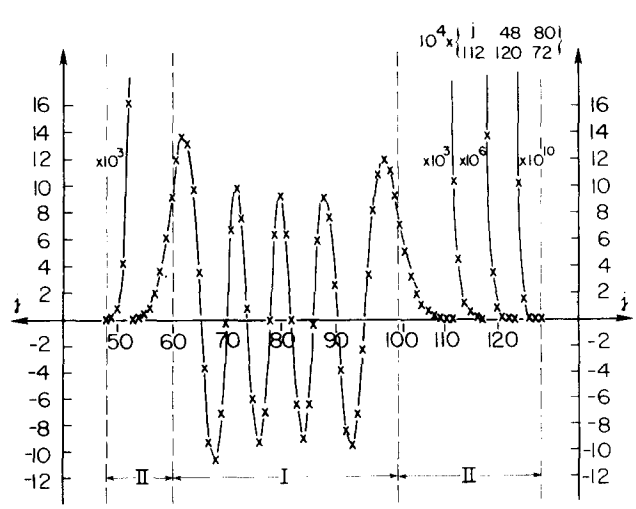


FIG. 3. Functional behavior of the $6j$-coefficients $h(j_1)$ $=\begin{Bmatrix} j_1 & 48 & 80 \\ 112 & 120 & 72 \end{Bmatrix}$ $(48 \le j_1 \le 128)$. The evaluation of $h(j_1)$ followed the recursion algorithm described in Sec. 3. The domain of $h(j_1)$ can be divided into a classical and two nonclassical regions (Ref. 4). In the classical region $h(j_1)$ oscillates with slowly varying amplitude; in the nonclassical regions $|h(j_1)|$ monotonically decays to zero.

cursive evaluation of all $f(j_1)$ according to

$$j_1 A(j_1+1) f(j_1+1) + B(j_1) f(j_1) + (j_1+1) A(j_1) f(j_1-1) = 0 \tag{6'}$$

can be performed. But, one should note that such a recursion procedure to generate the quantities $f(j_1)$, $f(j_1+1)$, $f(j_1+2)$, $\cdots$ can be numerically stable only in the direction of *increasing* $f(j_1)$. The semiclassical expressions for $3j$-coefficients,[4] reveal that $f(j_1)$ decreases rapidly to zero at the boundaries of the j_1-domain $j_{1\min}$ and $j_{1\max}$. This can also be seen from Fig. 1 which illustrates the typical j_1-dependence of $3j$-coefficients. In order to assure numerical stability, the recursive evaluation should therefore proceed from the boundaries $j_{1\min}$(left recursion) and $j_{1\max}$(right recursion) of the j_1-domain towards the middle (classical[4,13]) region. The classical region is defined here as the set of j_1-values for which there exists a classical angular momentum vector diagram corresponding to the $3j$-coefficient $f(j_1)$. It is within this region that the typical magnitudes of the $3j$-coefficients $f(j_1)$ are largest.

For the start of the recursion (6′) one observes that $A(j_{1\min})=0$ and $A(j_{1\max}+1)=0$. The recursion relation at the boundaries $j_{1\min}$ and $j_{1\max}$ thus becomes

$$B(j_{1\min}) f(j_{1\min}) + j_{1\min} A(j_{1\min}+1) f(j_{1\min}+1) = 0 \tag{16}$$

and

$$B(j_{1\max}) f(j_{1\max}) + (j_{1\max}+1) A(j_{1\max}) f(j_{1\max}-1) = 0, \tag{17}$$

i.e., the three term recursion (6′) reduces to two terms. Thus, *one* starting value at each boundary, namely $f(j_{1\min})$ and $f(j_{1\max})$, is sufficient to start the recursion (6′) in each direction.

Let us now assume that the terminal $6j$-coefficient $f(j_{1\min})$ and $f(j_{1\max})$ have been given arbitrary values and used to start the recursion (6′). Thus, they are in error by factors c_1 and c_2, respectively. Applications of Eqs. (16) and (17) then yield the quantities $c_1 f(j_{1\min})$ and $c_2 f(j_{1\max})$. Carrying the recursion further towards the classical regions by means of the linear recursion (6′), the quantities

$c_1 f(j_{1\min})$; $c_1 f(j_{1\min}+1)$; $\ldots$; $c_1 f(j_{1\mathrm{mid}})$ (left recursion),

$c_2 f(j_{1\max})$; $c_2 f(j_{1\max}-1)$; $\ldots$; $c_2 f(j_{1\mathrm{mid}})$ (right recursion)

will be generated. The common final j_1-value $j_{1\mathrm{mid}}$ for the recursions from left and right should lie within the classical j_1-domain. The recursions from the left and from the right must, however, match at $j_1 = j_{1\mathrm{mid}}$, so that we have the condition $c_1 f(j_{1\mathrm{mid}}) = c_2 f(j_{1\mathrm{mid}})$. We may therefore rescale the left recursion by the factor $c_2 f(j_{1\mathrm{mid}})/c_1 f(j_{1\mathrm{mid}}) = c_2/c_1$ to get

$$c_2 f(j_{1\min});\ c_2 f(j_{1\min}+1);\ \ldots;\ c_2 f(j_{1\max}-1)\ ;\ c_2 f(j_{1\max}), \tag{18}$$

i.e., the series of $3j$-coefficients in (1) off by a common factor c_2. To obtain the unknown c_2, we employ the normalization condition (7) which yields the absolute magnitude of c_2. The phase convention

$$\operatorname{sgn}\left\{\begin{pmatrix} j_{1\max} & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}\right\} = (-1)^{j_2-j_3-m_1} \tag{19}$$

determines the sign of c_2. Rescaling the series (18) by $1/c_2$ then gives the $6j$-coefficients in (1). It has, hence, been shown that the recursion (6′) can be started with arbitrarily chosen values $c_1 f(j_{1\min})$ and $c_2 f(j_{1\max})$ to obtain simultaneously all $3j$-coefficients in (1).

Let us consider now the evaluation of the $3j$-coefficients in (2),

$$g(m_2) = \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & -m_1-m_2 \end{pmatrix}, \quad m_{2\min} \le m_2 \le m_{2\max}, \tag{2'}$$

by means of the recursion equation

$$C(m_2+1) g(m_2+1) + D(m_2) g(m_2) + C(m_2) g(m_2-1) = 0. \tag{9'}$$

The range of allowed m_2-values in (2) is finite, the smallest m_2-value is $m_{2\min} = \max\{-j_2, -j_3-m_1\}$ and the largest m_2-value is $m_{2\max} = \min\{j_2, j_3-m_1\}$. The functional behavior of $g(m_2)$ resembles that of $f(j_1)$ in that $g(m_2)$ in general falls off to zero at the boundaries $m_{2\min}$ and $m_{2\max}$ of the m_2-domain (see also Fig. 2). To assure numerical stability, it is necessary to perform the recursion (9′) from both ends of the m_2-domain (left and right recursion). As was the case for (6′) the terminal recursions contain only the two terms

$$D(m_{2\min}) g(m_{2\min}) + C(m_{2\min}+1) g(m_{2\min}+1) = 0, \tag{20}$$

$$D(m_{2\max}) g(m_{2\max}) + C(m_{2\max}) g(m_{2\max}-1) = 0, \tag{21}$$

since $C(m_{2\min})=0$ and $C(m_{2\max}+1)=0$. Assuming arbitrary starting values $c_1 g(m_{2\min})$ and $c_2 g(m_{2\max})$, the recursion by means of (20), (9′), and (21), (9′) yields the two series

$c_1 g(m_{2\min})$; $c_1 g(m_{2\min}+1)$; $\ldots$;

$c_1 (m_{2\mathrm{mid}})$ (left recursion)

and

$c_2 g(m_{2\max})$; $c_2 g(m_{2\max}-1)$; $\ldots$;

$c_2 g(m_{2\mathrm{mid}})$ (right recursion)

for some $m_{2\mathrm{mid}}$ in the classical m_2-domain of the $3j$-coefficients. To assure that the matching condition $c_1 g(m_{2\mathrm{mid}}) = c_2 g(m_{2\mathrm{mid}})$ hold, the left recursion may be rescaled by the factor $c_2 g(m_{2\mathrm{mid}})/c_1 g(m_{2\mathrm{mid}})$. One then gets

$$c_2 g(m_{2\min});\ c_2 g(m_{2\min}+1);\ \ldots;$$

$$c_2 g(m_{2\max}-1);\ c_2 g(m_{2\max}) \tag{22}$$

which represents the $3j$-coefficients in (2), scaled by the unknown factor c_2. c_2 is readily determined from the

normalization condition (10) together with the phase convention

$$\operatorname{sgn}\left\{\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_{2\max} & -m_1-m_{2\max}\end{pmatrix}\right\} = (-1)^{j_2-j_3-m_1}. \tag{23}$$

The desired $3j$-coefficients are then obtained after multiplying (22) by $1/c_2$.

Finally, we turn to the recursive evaluation of the series of the $6j$-coefficients in (3)

$$h(j_1)=\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3\end{Bmatrix}, \quad j_{1\min}\leq j_1 \leq j_{1\max}. \tag{3'}$$

The smallest and largest j_1-values are $j_{1\min}=\max\{|j_2-j_3|,\ |l_2-l_3|\}$, and $j_{1\max}=\min\{j_2+j_3,\ l_2+l_3\}$. The $6j$-coefficients $h(j_1)$ fall off to zero at the boundaries $j_{1\min}$ and $j_{1\max}$ as can be seen from the example given in Fig. 3 and is revealed for the general case by the semiclassical expression for $6j$-coefficients.[4,11] Hence, the recursion

$$j_1 E(j_1+1)h(j_1+1)+F(j_1)h(j_1)+(j_1+1)E(j_1)h(j_1-1)=0 \tag{12'}$$

which connects all possible $h(j_1)$ should again proceed simultaneously from the boundaries $j_{1\min}$ (left recursion) and $j_{1\max}$ (right recursion) towards the middle j_1-domain. For the recursions at the boundaries we have $E(j_{1\min})=0$ and $E(j_{1\max}+1)=0$. Hence

$$F(j_{1\min})h(j_{1\min})+j_{1\min}E(j_{1\min}+1)h(j_{1\min}+1)=0 \tag{24}$$

and

$$F(j_{1\max})h(j_{1\max})+(j_{1\max}+1)E(j_{1\max})h(j_{1\max}-1)=0 \tag{25}$$

can be generated recursively for $j_{1\mathrm{mid}}$ being chosen to lie within the classical[4,13] j_1-domain. The classical domain of $6j$-coefficients is the domain of all quantum numbers for which there exists a classical angular momentum vector tetrahedron corresponding to the $6j$-coefficient. Within this domain, typical magnitudes of the $6j$-coefficients are largest. The matching condition $c_1 h(j_{1\mathrm{mid}})=c_2 h(j_{1\mathrm{mid}})$ is satisfied if the left recursion is rescaled by the factor $c_2 h(j_{1\mathrm{mid}})/c_1 h(j_{1\mathrm{mid}})$ which gives

$$c_2 h(j_{1\min});\ c_2 h(j_{1\min}+1);\ \ldots;\ c_2 h(j_{1\max}-1);\ c_2 h(j_{1\max}).$$

c_2 is determined from the normalization condition (13') together with the phase convention

$$\operatorname{sgn}\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3\end{Bmatrix} = (-1)^{j_2+j_3+l_2+l_3}, \tag{27}$$

so that finally all $6j$-coefficients in (3) are evaluated.

IV. ACCURACY AND EFFICIENCY OF RECURSIVE ALGORITHM

We would like to demonstrate now our claim that the recursive algorithm for the evaluation of $3j$- and $6j$-coefficients is numerically accurate for small and large quantum numbers and, in general, more efficient than existing algorithms based on the explicit expressions for these coefficients given by Wigner and Racah. As far as numerical effort is concerned, the advantageous character of a recursive evaluation is quite obvious. To obtain the coupling coefficients in (1), (2), and (3), essentially only the series $A(n)$, $B(n)$ or $C(n)$, $D(n)$ or $E(n)$, $F(n)$, respectively, which enter as coefficients the recursion equations (6), (9), and (12), need to be calculated.

The fact that the recursive algorithm evaluates a whole set of coupling coefficients is often an advantage, for in many problems of angular momentum coupling whole sets of coupling coefficients like (1), (2), or (3) enter. To give an example we may turn to the evaluation of $9j$-coefficients, which are given through the expansion

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \\ j_7 & j_8 & j_9\end{Bmatrix} = \sum_H (-1)^{2H} \begin{Bmatrix} H & j_1 & j_9 \\ j_3 & j_6 & j_2\end{Bmatrix} \begin{Bmatrix} H & j_2 & j_8 \\ j_5 & j_4 & j_8\end{Bmatrix} \begin{Bmatrix} H & j_9 & j_1 \\ j_7 & j_4 & j_8\end{Bmatrix}. \tag{28}$$

Evidently three strings of $6j$-coefficients are needed in the course of evaluating this expansion, namely

$$\begin{Bmatrix} H & j_1 & j_9 \\ j_3 & j_6 & j_2\end{Bmatrix}, \tag{3a}$$

$$\begin{Bmatrix} H & j_2 & j_8 \\ j_5 & j_4 & j_8\end{Bmatrix}, \tag{3b}$$

$$\begin{Bmatrix} H & j_9 & j_1 \\ j_7 & j_4 & j_8\end{Bmatrix}. \tag{3c}$$

Furthermore, to obtain the N $9j$-coefficients for all allowed j_3-quantum numbers, it is sufficient to evaluate (3b) and (3c) once and (3a) for all N allowed j_3 values. Hence, to determine the values of $9j$-coefficients for all j_3, only $N+2$ $6j$-recursions have to be performed. These considerations exemplify how strings of coupling coefficients like (1), (2), and (3) naturally enter into the problems of angular momentum coupling.

To answer the important question about the numerical accuracy of the proposed algorithm, a comparison between recursively evaluated coupling coefficients and tabulated values of these coefficients suggests itself. The exact representation of $3j$- and $6j$-coefficients in terms of prime number factors, as given in the table of Rotenberg *et al.*[14] provides the accurate values for these coefficients. In Tables I, II, and III we present a comparison between $3j$- and $6j$-coefficients generated by recursion and those obtained from the tabulation of Rotenberg *et al.* As can be seen, agreement is found for essentially all significant figures provided by the computer representation of numerical constants (i.e., 16 significant digits in the double precision mode on a IBM 360/91).

Perhaps more important is the fact that the recursive algorithm allows the evaluation of coupling coefficients with very large quantum numbers, thus enlarging the realm of coupling coefficients accessible to numerical methods. Since no tabulated values of large quantum number coupling coefficients exist, the accuracy of the recursive algorithm must be demonstrated through a test of its numerical stability. This has been done by carrying out two simultaneous evaluations of $6j$-coefficients for large quantum numbers, the results of which

TABLE I. Accuracy of recursively evaluated $3j$-coefficients $\begin{pmatrix} L1 & 9/2 & 7/2 \\ 1 & -7/2 & 5/2 \end{pmatrix}$.

$L1$	Values of $3j$-coefficients[a]
1	0.278 886 675 511 3585 (0) I 0.278 886 675 511 3586 (0) II
2	−0.953 462 589 245 5920 (−1) I −9.953 462 589 245 5920 (−1) II
3	−0.674 199 862 463 2420 (−1) I −0.674 199 862 463 2420 (−1) II
4	0.153 311 035 167 9666 (0) I 0.153 311 035 167 9666 (0) II
5	−0.156 446 554 693 6860 (0) I −0.156 446 554 693 6859 (0) II
6	0.109 945 041 215 6550 (0) I 0.109 945 041 215 6550 (0) II
7	−0.553 623 569 313 1718 (−1) I −0.553 623 569 313 1718 (−1) II
8	0.179 983 545 113 7785 (−1) I 0.179 983 545 113 7785 (−1) II

[a] I: Rotenberg *et al.*; II: This paper.

TABLE III. Accuracy of recursively evaluated $6j$-coefficients $\begin{Bmatrix} L1 & 8 & 7 \\ 13/2 & 15/2 & 15/2 \end{Bmatrix}$.

$L1$	Values of $6j$-coefficients[a]
1	0.349 090 513 837 3299 (−1) I 0.349 090 513 837 3284 (−1) II
2	−0.374 302 503 965 9791 (−1) I −0.374 302 503 965 9775 (−1) II
3	0.189 086 639 095 9559 (−1) I 0.189 086 639 095 9551 (−1) II
4	0.734 244 825 492 8642 (−2) I 0.734 244 825 492 8610 (−2) II
5	−0.235 893 518 508 1794 (−1) I −0.235 893 518 508 1783 (−1) II
6	0.191 347 695 521 5436 (−1) I 0.191 347 695 521 5427 (−1) II
7	0.128 801 739 772 4172 (−2) I 0.128 801 739 772 4175 (−2) II
8	−0.193 001 836 629 0526 (−1) I −0.193 001 836 629 0531 (−1) II

[a] I: Rotenberg *et al.*; II. This paper.

are presented in Table IV. One of the calculations was done in single precision mode (IBM 360/91) which provides 6 significant digits for numerical constants, and the other calculations used double precision with 16 significant digits. Numerical stability is demonstrated since the single precision calculation agrees with the double precision calculation within its full range of significant figures, as is shown for one example in Table IV. It is remarkable that even the small coefficients near the ends of the range are found with the maximum possible relative accuracy.

TABLE II. Accuracy of recursively evaluated $3j$-coefficients $\begin{pmatrix} 8 & 15/2 & 13/2 \\ 1 & M & -M-1 \end{pmatrix}$.

M	Values of $3j$-coefficients[a]
−15/2	0.209 158 973 288 6152 (−1) I 0.209 158 973 288 6155 (−1) II
−13/2	0.853 756 555 321 5250 (−1) I 0.853 756 555 321 5260 (−1) II
−11/2	0.908 295 370 868 6920 (−1) I 0.908 295 370 868 6930 (−1) II
−9/2	−0.389 054 377 846 4993 (−1) I −0.389 054 377 846 4998 (−1) II
−7/2	−0.663 734 970 165 6800 (−1) I −0.663 734 970 165 6810 (−1) II
−5/2	−0.649 524 040 528 3890 (−1) I −0.649 524 040 528 3900 (−1) II
−3/2	−0.215 894 310 595 4037 (−1) I −0.215 894 310 595 4036 (−1) II
−1/2	−0.778 912 711 785 2390 (−1) I −0.778 912 711 785 2390 (−1) II
1/2	0.359 764 371 059 5433 (−1) I 0.359 764 371 059 5431 (−1) II
3/2	0.547 301 500 021 2632 (−1) I 0.547 301 500 021 2631 (−1) II
5/2	−0.759 678 665 956 7610 (−1) I −0.759 678 665 956 7610 (−1) II
7/2	−0.219 224 445 539 8920 (−1) I −0.219 224 445 539 8921 (−1) II
9/2	0.101 167 744 280 7722 (0) I 0.101 167 744 280 7721 (0) II
11/2	0.734 825 726 244 7190 (−1) I 0.734 825 726 244 7190 (−1) II

[a] Rotenberg *et al.*; II: This work.

APPENDIX: DERIVATION OF RECURSION EQUATION FOR $6j$-COEFFICIENTS AS SOLUTIONS TO AN EIGENVALUE PROBLEM

The remarkable resemblance between the series of

TABLE IV. Accuracy of recursively evaluated $6j$-coefficients $\begin{Bmatrix} L1 & 120 & 72 \\ 112 & 48 & 80 \end{Bmatrix}$.

$L1$	Values of $6j$-coefficients[a]
48	0.161 1825 (−8) I 0.161 1825 (−8) II
56	0.949 0981 (−4) I 0.949 0977 (−4) II
64	0.964 1119 (−3) I 0.964 1123 (−3) II
72	0.988 0543 (−3) I 0.988 0543 (−3) II
80	0.919 9395 (−3) I 0.919 9396 (−3) II
88	0.918 3130 (−3) I 0.918 3132 (−3) II
96	0.332 5452 (−3) I 0.332 5446 (−3) II
104	0.322 0624 (−3) I 0.322 0624 (−3) II
112	0.115 3951 (−5) I 0.115 3951 (−5) II
120	0.835 0757 (−10) I 0.835 0761 (−10) II
128	0.119 3770 (−16) I 0.119 3771 (−16) II

[a] I: double precision; II: single precision.

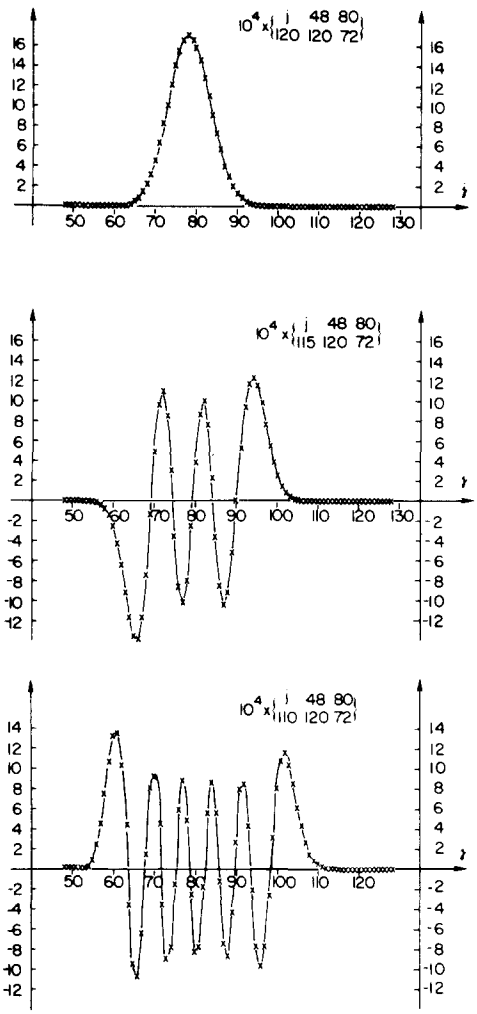


FIG. 4. Comparison of the functional behavior of the $6j$-coefficients $h(j_1) = \begin{Bmatrix} j_1 & 48 & 80 \\ l_1 & 120 & 72 \end{Bmatrix}$ $(48 \leq j_1 \leq 128)$ evaluated through the recursion algorithm described in Sec. 3. The largest allowed l_1-quantum number is $l_{1\max} = 120$. The diagram shows that $l_{1\max} - l_1$ counts the nodes of $h(j_1)$: (a) $l_{1\max} - l_1 = 0$ (b) $l_{1\max} - l_1 = 5$; (c) $l_{1\max} - l_1 = 10$.

$3j$- and $6j$-coefficients in Figs. 1, 2, 3, and bound state eigenfunctions may not have escaped the readers' attention. To carry further a comparison between angular momentum coupling coefficients and eigensolutions of bound state problems we present in Fig. 4 the series of $6j$-coefficients

$$[(2j_1+1)(2l_1+1)]^{1/2} \begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix}, \quad j_{1\min} \leq j_1 \leq j_{1\max},$$

for different l_1-values. $j_{1\min} = \max\{|j_2 - j_3|, |l_2 - l_3|\}$ and $j_{1\max} = \min\{j_2 + j_3, l_2 + l_3\}$ are the smallest and largest values j_1 can assume in order for the $6j$-coefficients not to vanish. It can be seen from Fig. 4 that l_1 takes on the character of a quantum number which counts the nodes $l_{1\max} - l_1$ of the series $(l_{1\max} = \min\{j_2 + l_3,\ j_3 + l_3\})$. What is the origin of this particular behavior of the coupling coefficients? The answer to this question is that $3j$- and $6j$-coefficients are by definition components of eigenvectors to certain eigenvalue problems. The coupling coefficients in Figs. 1–4 just represent those eigenvectors.

That $3j$-coefficients can be obtained through the diagonalization of certain angular momentum operators has been known since the early days of quantum mechanics. Hence, this will not be demonstrated here, but we may refer the reader to Refs. 5 and 6. However, we will show in the following which eigenvalue problem defines $6j$-coefficients, and will prove that the recursion equations for $6j$-coefficients are a consequence of this eigenvalue problem. The main reason for this algebraic detour is to convince the reader that the recursion equations derived above do indeed follow directly from the definition of the 6-j coefficients.

Let us consider a system composed of four angular momenta $\mathbf{J}_2, \mathbf{J}_3, \mathbf{L}_2$, and $\mathbf{L}_3$ such that $\mathbf{J}_2 + \mathbf{J}_3 + \mathbf{L}_2 + \mathbf{L}_3 = 0$. This system may be described by two different *zero* total angular momentum states:

$$|(j_2, j_3)(l_2, l_3)j_1\rangle = \sum_m (-1)^{j_1 - m} [2j_1 + 1]^{-1/2} |(j_2, j_3)j_1\, m\rangle\, |(l_2, l_3)j_1 - m\rangle \tag{A1}$$

and

$$|(j_2, l_3)(l_2, j_3)l_1\rangle = \sum_m (-1)^{l_1 - m} [2l_1 + 1]^{-1/2} \times |(j_2, l_3)l_1 m\rangle\, |(l_2, j_3)l_1 - m\rangle. \tag{A2}$$

The transformation matrix element $\langle (j_2, j_3)(l_2, l_3)j_1 | \times (j_2, l_3)(l_2, j_3)l_1\rangle$ defines then the $6j$-coefficient

$$\langle (j_2, j_3)(l_2, l_3)j_1 | (j_2, l_3)(l_2, j_3)l_1\rangle = [(2j_1+1)(2l_1+1)]^{1/2} \begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix}. \tag{A3}$$

It is a simple exercise in angular momentum algebra to show that this definition is in agreement with the more conventional definition of $6j$-coefficients in terms of $3j$-coefficients.[10] $|(j_2, j_3)(l_2, l_3)j_1\rangle$ and $|(j_2, l_3)(l_2, j_3)l_1\rangle$ are both eigenstates of the angular momentum operators J_2^2, J_3^2, L_2^2, L_3^2, but only the first state is also an eigenfunction of $J_1^2 = (\mathbf{J}_2 + \mathbf{J}_3)^2$, whereas only the latter is an eigenstate of $L_1^2 = (\mathbf{J}_2 + \mathbf{L}_3)^2$. From elementary principles of linear algebra it then follows that the columns of $\langle (j_2, j_3)(l_2, l_3)j_1 | (j_2, l_3)(l_2, j_3)l_1\rangle$ are the eigenvectors of the operator L_1^2 in the $|(j_2, j_3)(l_2, l_3)j_1\rangle$-basis.

Let us evaluate L_1^2 in this basis. We first notice that $[\mathbf{J}_2 + \mathbf{J}_3 + \mathbf{L}_2 + \mathbf{L}_3,\ L_1^2] = 0$, hence, the application of L_1^2 does not affect the total angular momentum state. We may then express L_1^2 through operators whose action on the intermediate states $|(j_2, j_3)j_1 m\rangle$ and $|(l_2, l_3)j_1 - m\rangle$ in (1) are known:

$$L_1^2 = J_2^2 + L_3^2 + J_{2-}L_{3+} + J_{2+}L_{3-} + 2J_{2z}L_{3z}. \tag{A4}$$

We have obviously

$$(J_2^2 + L_3^2)\, |(j_2, j_3)(l_2, l_3)j_1\rangle = [j_2(j_2+1) + l_3(l_3+1)]\, |(j_2, j_3)(l_2, l_3)j_1\rangle.$$

The remaining operators in (4) applied to $|(j_2, j_3)(l_2, l_3)j_1\rangle$ give

$$\sum_m \frac{(-1)^{j_1 - m}}{[2j_1+1]^{1/2}} \Big\{ \langle (j_2, j_3)j_1 m\, | J_{2-}\, \langle (l_2, l_3)j_1 - m\, | L_{3+}$$

$$+\langle(j_2,j_3)j_1m\,|J_{2+}\langle(l_2,l_3)j_1-m\,|L_{3-}$$

$$+2\langle(j_2,j_3)j_1m\,|J_{2z}\langle(l_2,l_3)j_1-m\,|L_{3z}\Big\}. \quad \text{(A5)}$$

The operators $J_{2-}L_{3+}, J_{2+}L_{3-}$ and $J_{2z}L_{3z}$ only couple to states $|(j_2,j_3)j_1'm+1\rangle|(l_2,l_3)j_1''-m-1\rangle)$, $|(j_2,j_3)j_1'm-1\rangle \times|(l_2,l_3)j_1''-m+1\rangle$, and $|(j_2,j_3)j_1'm\rangle|(l_2,l_3)j_1''-m\rangle$ where $j_1' = j_1\pm1$, 0 and $j_1''=j_1\pm1$, 0.[15] But, in order that the state (5) carries zero total angular momentum all terms with $j_1'\neq j_1''$ must cancel out, and hence may be disregarded in the following calculation. The matrix elements $\langle(j_2, j_3)j_1m\,|J_{2-}|(j_2,j_3)j_1'm'\rangle$, $\langle(l_2,l_3)j_1-m\,|L_{3+}|(l_2,l_3)j_1'-m'\rangle$, etc. all split in orientation independent factors $\langle j_1\|J_2\|j_1'\rangle$ and $\langle j_1\|L_3\|\ j_1'\rangle$ and m-dependent factors (Wigner–Eckart theorem). The m-dependent factors may be obtained from Ref. 15. It should further be noted that the operators L_{3+}, L_{3z} operate on the second angular momentum l_3 in $|(l_2,l_3)j_1-m\rangle$, whereas the operators J_{2+}, J_{2z} operate on the first angular momentum j_2 in $|(j_2,j_3)j_1m\rangle$. This makes it necessary[5] to give negative values to the off-diagonal elements $\langle j_1\|L_3\|j_1\pm1\rangle$. We obtain then from (5) the expression

$$\sum_m \frac{(-1)^{j_1-m}}{[2j_1+1]^{1/2}}\{-\langle j_1-1\|J_2\|j_1\rangle\langle j_1-1\|L_3\|j_1\rangle$$

$$\times[-(j_1-m-1)(j_1-m)\langle(j_2,j_3)j_1-1\ m+1\,|\langle(l_2,l_3)j_1-1\ -m-1\,|$$

$$-(j_1+m)(j_1+m-1)\langle(j_2,j_3)j_1-1\ \ m-1\,|\langle(l_2,l_3)j_1-1\ -m+1\,|$$

$$+2(j_1^2-m^2)\langle(j_2,j_3)j_1-1\ \ m\ |\langle(l_2,l_3)j_1-1\ -m\,|]$$

$$+\langle j_1\|J_2\|j_1\rangle\langle j_1\|L_3\|j_1\rangle$$

$$\times[(j_1+m+1)(j_1-m)\langle(j_2,j_3)j_1m+1\,|\langle(l_2,l_3)j_1\ -m-1\,|$$

$$+(j_1+m)(j_1-m+1)\langle(j_2,j_3)j_1\ m-1\,|\langle(l_2,l_3)j_1\ -m+1\,|$$

$$-2m^2\ \langle(j_2,j_3)j_1\ \ m\,|\langle(l_2,l_3)j_1\ -m\,|]$$

$$-\langle j_1+1\|J_2\|j_1\rangle\langle j_1+1\|L_3\|j_1\rangle$$

$$\times[-(j_1+m+1)(j_1+m+2)\langle(j_2,j_3)j_1+1\ m+1\,|$$

$$\times\langle(l_2,l_3)j_1+1\ -m-1\,|$$

$$-(j_1-m+1)(j_1-m+2)\ \langle(j_2,j_3)j_1+1\ m-1\,|$$

$$\times\langle(l_2,l_3)j_1+1\ -m+1\,|$$

$$+2[(j_1+1)^2-m^2]\ \langle(j_2,j_3)j_1+1\ m\ |\langle(l_2,l_3)j_1+1\ -m\,|]\}. \quad \text{(A6)}$$

Collecting terms with equal magnetic quantum numbers leads to the cancellation of all m-dependent prefactors in the sum, except for the phase $(-1)^{-m}$. Carrying out the m-summation gives then

$$2j_1[(2j_1+1)(2j_1-1)]^{1/2}\langle j_1-1\|J_2\|j_1\rangle\langle j_1-1\|L_3\|j_1\rangle$$

$$\times\langle(j_2,j_3)(l_2,l_3)j_1-1\,|$$

$$-2j_1(j_1+1)\langle j_1\|J_2\|j_1\rangle\langle j_1\|L_3\|j_1\rangle\langle(j_2,j_3)(l_2,l_3)j_1\,|$$

$$+2(j_1+1)[(2j_1+1)(2j_1+3)]^{1/2}$$

$$\times\langle j_1+1\|J_2\|j_1\rangle\langle j_1+1\|L_3\|j_1\rangle\langle(j_2,j_3)(l_2,l_3)j_1+1\,| \quad \text{(A7)}$$

and, finally, with the explicit algebraic expression for $\langle j_1\|J_2\|j_1'\rangle$ and $\langle j_1\|L_3\|j_1'\rangle$[5,15]

$$\langle j_1|L_1^2|j_1-1\rangle = \frac{\{[j_1^2-(j_2-j_3)^2][(j_2+j_3+1)^2-j_1^2][j_1^2-(l_2-l_3)^2][(l_2+l_3+1)-j_1^2]\}^{1/2}}{2j_1[(2j_1-1)(2j_1+1)]^{1/2}}, \quad \text{(A8)}$$

$$\langle j_1|L_1^2|j_1+1\rangle = \frac{\{[(j_1+1)^2-(j_2-j_3)^2][(j_2+j_3+1)^2-(j_1+1)^2][(j_1+1)^2-(l_2-l_3)^2][(l_2+l_3+1)^2-(j_1+1)^2]\}^{1/2}}{2(j_1+1)[(2j_1+1)(2j_1+3)]^{1/2}}, \quad \text{(A9)}$$

$$\langle j_1|L_1^2|j_1\rangle = \{j_1(j_1+1)[-j_1(j_1+1)+j_2(j_2+1)+j_3(j_3+1)]+l_2(l_2+1)[j_1(j_1+1)+j_2(j_2+1)-j_3(j_3+1)]+l_3(l_3+1) \quad \text{(A10)}$$

$$\times[j_1(j_1+1)-j_2(j_2+1)+j_3(j_3+1)]\}[2j_1(j_1+1)]^{-1}$$

These matrix elements show that L_1^2 is a symmetric, tridiagonal matrix. L_1^2 is readily diagonalized and must have real, positive eigenvalues. However, such a diagonalization procedure would provide redundant results (eigenvectors and eigenvalues), since the eigenvalues of L_1^2 are known to be $l_1(l_1+1)$ with $l_1=l_{1\,\max}-n$, $n=0,1,2,\ldots$. Hence, it is sufficient to solve the system of homogeneous equations

$$[L_1^2-l_1(l_1+1)]\mathbf{x}=0. \quad \text{(A11)}$$

Because of the tridiagonal form of L_1^2, this leads to three-term recursion equations for the components of $\mathbf{x}$. These equations are identical with the recursion equations derived above for $6j$-coefficients. The solution of these recursion equations therefore corresponds directly to the solution of the eigenvalue problem (11).

*Supported by the National Science Foundation.

†Current address: Max Planck Institut für biophysikalische Chemie, Göttingen, Germany.

[1]E. P. Wigner, *Group Theory* (Academic, New York, 1959).
[2]G. Racah, Phys. Rev. 62, 438 (1942).
[3]K. Schulten and R. G. Gordon (to be published).
[4]K. Schulten and R. G. Gordon, J. Math. Phys. 15, 1971 (1975), following paper.
[5]E. U. Condon and G. H. Shortley, *The Theory of Atomic*

Spectra (Cambridge U. P., Cambridge, 1935), p. 67.
[6]M. E. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1957), pp. 42—45.
[7]A. P. Yutsis, I. B. Levinson, and V. V. Vanagas, "The Mathematical Apparatus of the Theory of Angular Momentum"; Israel Program for Scientific Translation (Jerusalem, 1962).
[8]A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1965), Vol. II.
[9]J. D. Louck, Phys. Rev. 110, 815 (1958).
[10]A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, 1957).
[11]U. Fano and G. Racah, *Irreducible Tensorial Sets* (Academic, New York, 1959), Appendix.
[12]K. Alder, A. Bohr, T. Huus, B. Mottelson, and A. Winther, Rev. Mod. Phys. 28, 432 (1956).
[13]G. Ponzano and T. Regge, in *Spectroscopic and Group Theoretical Methods in Physics* (North-Holland, Amsterdam, 1968).
[14]M. Rotenberg, R. Bivins, N. Metropolis, and J. K. Wooten, Jr., *The 3-j and 6-j Symbols* (MIT Press, Cambridge, Mass., 1959).
[15]P. Güttinger and W. Pauli, Z. Phys. 67, 743 (1931).

Semiclassical approximations to 3*j*- and 6*j*-coefficients for quantum-mechanical coupling of angular momenta*

Klaus Schulten† and Roy G. Gordon

Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138
(Received 21 April 1975)

The coupling of angular momenta is studied using quantum mechanics in the limit of large quantum numbers (semiclassical limit). Uniformly valid semiclassical expressions are derived for the 3*j* (Wigner) coefficients coupling two angular momenta, and for the 6*j* (Racah) coefficients coupling three angular momenta. In three limiting cases our new expressions reduce to those conjectured by Ponzano and Regge. The derivation involves solving the recursion relations satisfied by these coefficients, by a discrete analog of the WKB method. Terms of the order of the inverse square of the quantum numbers are neglected in the derivation, so that the results should be increasingly accurate for larger angular momenta. Numerical results confirm this asymptotic convergence. Moreover, the results are of a useful accuracy even at small quantum numbers.

I. INTRODUCTION

3*j*- and 6*j*-coefficients describe the quantum mechanical coupling of two and three angular momentum states, respectively. Classically this coupling corresponds to the addition of angular momentum vectors. For larger angular momenta the classical concept of vector addition becomes increasingly valid; so that in the limit of very large quantum numbers 3*j*- and 6*j*-coefficients should have an interpretation in terms of classical vector diagrams (Figs. 1a, 2a, 3a). One can expect that in this semiclassical limit 3*j*- and 6*j*-coefficients can be expressed as simple functions of the geometric variables which describe the classical angular momentum addition. Such a functional relationship between angular momentum coupling coefficients and classical vector diagrams has been suggested, on the basis of heuristic arguments, by Ponzano and Regge.[1] It is one aim of this paper to give a rigorous derivation for the expressions of Ponzano and Regge as the asymptotic (WKB) solutions of the recursion equations which define 3*j*- and 6*j*-coefficients.

For certain quantum mechanically allowed 3*j*- and 6*j*-coefficients, classical vector diagrams do not exist. For example, the classical angular momenta associated with some 6*j*-coefficients cannot be connected to give a vector tetrahedron as in Fig. 3a, and such cases are called classically forbidden. Thus, the quantum number domains of the angular momentum coupling coefficients are to be separated into classically allowed and classically forbidden regions. Interestingly enough, the algebraic definition of the geometric variables (volume and dihedral angles of the 6*j*-tetrahedron, etc.) can be continued from the classical domain of quantum numbers to the nonclassical domains. Accordingly, Ponzano and Regge stated three different expressions for 3*j*- and 6*j*-coefficients valid in either the classical domain, the nonclassical domain or at the boundary between these two domains. These expressions do not, however, smoothly connect with each other, and lacking a systematic derivation of their results Ponzano and Regge were not able to correct this deficiency.

Miller,[2] starting from the correspondence relations of classical and quantum mechanics, recently derived a semiclassical expression for 3*j*-coefficients restricted to the classically allowed domain of quantum numbers. Miller's result is identical with the corresponding expression of Ponzano and Regge giving, thus, support to the supposition that Ponzano and Regge's formulas are amenable to a rigorous derivation. We have, in fact, found such derivation and present it here. The route we have taken for this derivation will be outlined now.

One may recall that semiclassical expressions for quantum mechanical wavefunctions can be determined as asymptotic solutions of the Schrödinger second order differential equation by means of the WKB approximation. The 3*j*- and 6*j*-coefficients are the solutions of certain linear recursion equations which, as we have shown,[3] provide the most efficient and stable algorithm for their evaluation even for very large quantum numbers. The important role of these recursion equations in determining, except for overall factors, the angular momentum coupling coefficients had been well known in the early days of quantum mechanics[4] but sank into oblivion after the advent of group theory and the following derivation of closed expressions of 3*j*- and 6*j*-coefficients by Wigner[5] and Racah.[6] The three term recursion equations can be formally written as second order *difference* equations which are closely related to the second order differential equations which result from the Schrödinger equation for wavefunctions. In view of this relationship, it seems reasonable to attempt a derivation of the semiclassical formulas for 3*j*- and 6*j*-coefficients as the asymptotic (WKB) solutions of their recursion equations. To this purpose we extended the WKB theory from differential to difference equations and this yields then the semiclassical expressions of Ponzano and Regge.

The recursion equations determine the angular momentum coupling coefficients over a discrete domain of quantum numbers. Classical angular momenta vary over a continuous domain of values. According to the correspondence of classical and quantum mechanics one expects that the semiclassical expressions of 3*j*- and 6*j*-coefficients can be defined over continuous domains also. In fact, we have found that the asymptotic solutions of the 3*j*- and 6*j*-recursion equations viewed

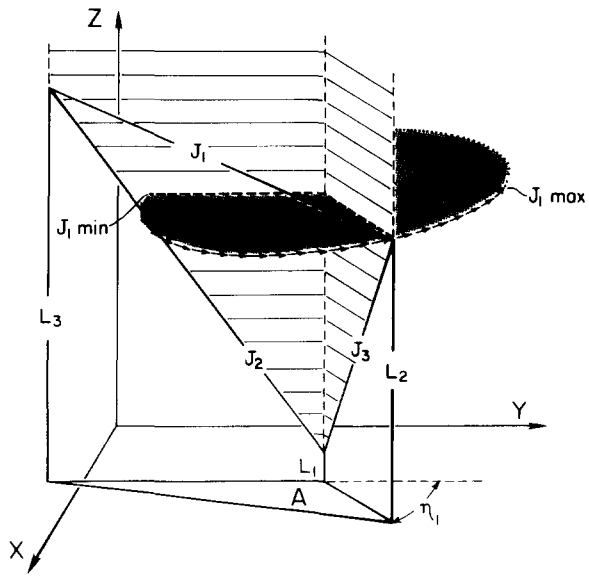
as functions of continuous variables obey asymptotically certain differential equations. Hence, the semiclassical $3j$- and $6j$-coefficients can also be understood as the WKB solutions of these second order differential equations in order to account for the continuous variation of the angular momenta in the semiclassical limit. If one employs the uniform WKB approximation to these differential equations, one obtains the $3j$- and $6j$-coefficients in terms of Airy function formulas which are uniformly valid over the entire domain of quantum numbers and, thus, present considerable improvement in accuracy, over the expressions given by Ponzano and Regge.

In Sec. 2 we will present the recursion equations of $3j$- and $6j$-coefficients, and show that in the semiclassical limit these equations are connected algebraically with classical vector diagrams. In Sec. 3 we derive the WKB solutions for second order difference equations, and demonstrate that these solutions also asymptotically satisfy associated differential equations. In Sec. 4 we apply the WKB approximation to obtain the semiclassical $3j$-coefficients in analytic form. In Sec. 5 these calculations are repeated to derive the semiclassical $6j$-coefficients. Finally, we compare in Sec. 6 the exact and the semiclassical values of $3j$- and $6j$-coefficients and demonstrate their convergence for large quantum numbers.

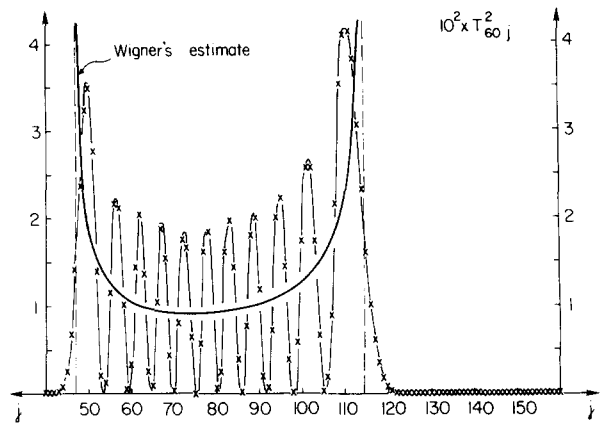
II. SEMICLASSICAL RECURSION EQUATIONS FOR $3j$-COEFFICIENTS

The Wigner $3j$-coefficients define the algebra of the quantum mechanical addition of angular momenta.[7] For a classical mechanical system with two internal angular momenta $\mathbf{J}_2$ and $\mathbf{J}_3$ the resulting total angular momentum is uniquely determined as $\mathbf{J}_1 = \mathbf{J}_2 + \mathbf{J}_3$. Quantum mechanically, it is not possible to specify for the angular momenta $\mathbf{J}_2$ and $\mathbf{J}_3$ simultaneously all three Cartesian vector components, but at most one component. If the z-components m_2 and m_3 are thought to be specified the system is said to be prepared in the internal angular momentum state $|j_2 m_2\rangle |j_3 m_3\rangle$. The relative orientation of the angular momenta $\mathbf{J}_2$ and $\mathbf{J}_3$ in the x,y plane, measured by an angle η_1 is then necessarily undetermined, each angle η_1 being equally likely. Hence, the total angular momentum $|\mathbf{J}_1|$ can assume a variety of values depending on the relative orientations η_1 of $\mathbf{J}_2$ and $\mathbf{J}_3$. This situation is depicted by the classical vector diagram in Fig. 1a.

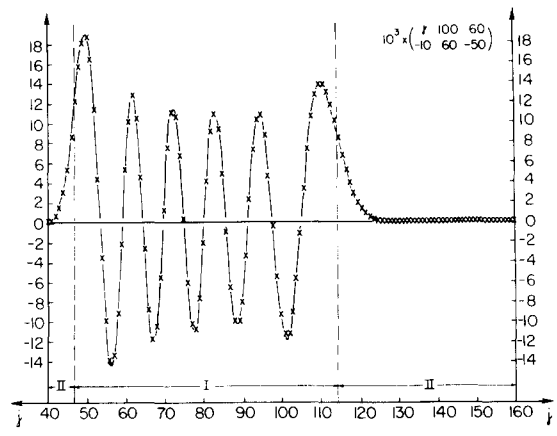
The system under consideration may also be prepared in a particular total angular momentum state $|(j_2,j_3)j_1 m_1\rangle$. For a system prepared in such a fashion the components of the angular momenta $\mathbf{J}_2$ and $\mathbf{J}_3$ have fixed projections along $\mathbf{J}_1$. The remaining components of $\mathbf{J}_2$ and $\mathbf{J}_3$, perpendicular to $\mathbf{J}_1$, are then necessarily undetermined. This situation is illustrated by the classical vector diagram in Fig. 2a. In particular, the z-components of $\mathbf{J}_2$ and $\mathbf{J}_3$ can take on a variety of values, depending on the orientation Θ_1 of $\mathbf{J}_2$ and $\mathbf{J}_3$ in a plane perpendicular to $\mathbf{J}_1$. Because of the randomness of the orientations of $\mathbf{J}_2$ and $\mathbf{J}_3$, each angle Θ_1 is equally likely.



(a)

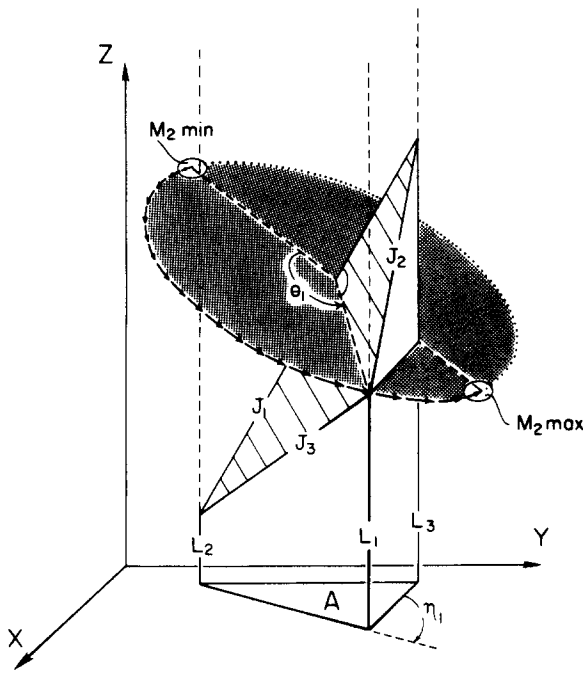


(b)

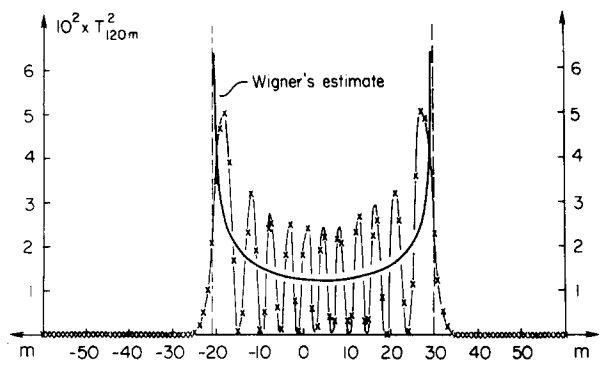


(c)

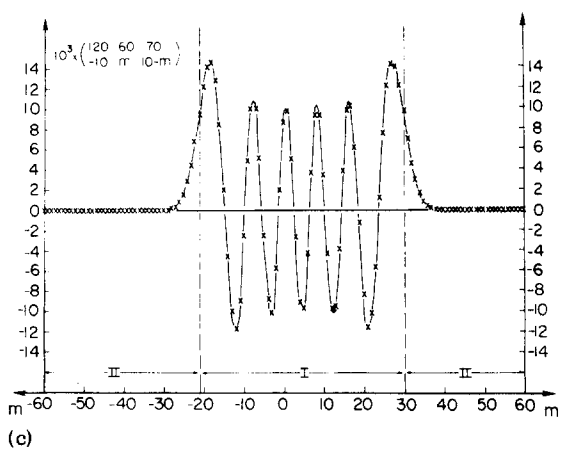
FIG. 1. The series of $3j$-coefficients $\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}$, $j_1 \min \leq j_1 \leq j_1 \max$, in (c) and the corresponding manifold of classical angular momentum vector diagrams $\mathbf{J}_1 + \mathbf{J}_2 + \mathbf{J}_3 = 0$ $(J_i = j_i + \frac{1}{2}, J_{iz} = m_i)$ generated by rotation around the shaded circle in (a). The quantum mechanical probability distribution $(2j_1+1)\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}^2$ for the occurrence of the classical vector diagrams in (a) are compared in (b) with Wigner's semiclassical estimate $(2j_1+1)/4\pi A$.



(a)



(b)



(c)

FIG. 2. The series of $3j$-coefficients $\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}$, $m_2\,\min \leq m_2 \leq m_2\,\max$, in (c) and the corresponding manifold of classical angular momentum vector diagrams $\mathbf{J}_1+\mathbf{J}_2+\mathbf{J}_3=0$ ($J_i=j_i+\frac{1}{2}$, $J_{iz}=m_i$) generated by rotation around the shaded circle in (a). The quantum mechanical probability distribution $(2j_1+1)\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}^2$ for the occurrence of the classical vector diagrams in (a) are compared in (b) with Wigner's semiclassical estimate $(2j_1+1)/4\pi A$.

The unitary transformation T between the representation $|j_2m_2\rangle\,|j_3m_3\rangle$ and $|(j_2,j_3)j_1m_1\rangle$ of the system of two angular momenta

$$|j_2m_2\rangle\,|j_3m_1-m_2\rangle=\sum_{j_1} T_{m_2j_1}\,|(j_2,j_3)j_1m_1\rangle \tag{1}$$

defines the $3j$-coefficients

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ -m_1 & m_2 & m_1-m_2 \end{pmatrix}=(-1)^{j_2-j_3+m_1}\,[2j_1+1]^{-1/2}\,T_{m_2j_1}. \tag{2}$$

T is customarily chosen real with the phase convention for its row vectors adopted as by Wigner. The squares $T_{m_2j_1}{}^2$ of the elements of T are to be interpreted as the probability for a system prepared in the internal angular momentum state $|j_2m_2\rangle\,|j_3m_1-m_2\rangle$ to be found in the total angular momentum state $|(j_2,j_3)j_1m_1\rangle$ and, vice versa, as the probability for a system prepared in the total angular momentum state $|(j_2,j_3)j_1m_1\rangle$ to be found in the internal angular momentum state $|j_2m_2\rangle\,|j_3m_1-m_2\rangle$. On the basis of this interpretation Wigner established an approximate functional expression for $3j$-coefficients, using the classical vector diagrams in Figs. 1a and 2a. The probability that in Fig. 1a $j_1 \leq |\mathbf{J}_2+\mathbf{J}_3| \leq j_1+1$, given by $T_{m_2j_1}{}^2$, can be evaluated from the fact that each relative orientation η_1 of $\mathbf{J}_2$ and $\mathbf{J}_3$ is equally likely. Wigner found[5]

$$T_{m_2j_1}{}^2=(2j_1+1)/4\pi A \tag{3}$$

where A, the area of the triangle $\Delta(\mathbf{J}_1,\mathbf{J}_2,\mathbf{J}_3)$ projected onto the x,y plane, is given by the Cayley determinant

$$A^2=-\frac{1}{16}\begin{vmatrix} 0 & J_1{}^2-m_1{}^2 & J_2{}^2-m_2{}^2 & 1 \\ J_1{}^2-m_1{}^2 & 0 & J_3{}^2-m_3{}^2 & 1 \\ J_2{}^2-m_2{}^2 & J_3{}^2-m_3{}^2 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{vmatrix}. \tag{4}$$

Likewise, the probability for $m_2 \leq J_{2z} \leq m_2+1$ in Fig. 2a also given by $T_{m_2j_1}{}^2$ can be evaluated by assuming that each angle Θ_1 is equally likely. The result is again (3).

In Figs. 1c and 2c the $3j$-coefficients $\begin{pmatrix} j_1 & 100 & 60 \\ -10 & 60 & -50 \end{pmatrix}$ and $\begin{pmatrix} 120 & 60 & 70 \\ -10 & m_2 & 10-m_2 \end{pmatrix}$ are plotted representing the quantum mechanical probability amplitudes corresponding to the classical angular momentum coupling depicted by Figs. 1a and 2a. Figures 1b and 2b present the associated probabilities $T_{m_2j_1}{}^2$ for a comparison with the Wigner expression (3). One can see that Eq. (3) does not really approximate individual $3j$-coefficients, but does provide an approximation for the average taken over a few neighboring $3j$-coefficients. Our aim now is to show how Wigner's estimate can be refined to give more accurate expressions for individual $3j$-coefficients.

The $3j$-coefficients in Figs. 1c and 2c are determined except for an overall constant factor as solutions of recursion equations. The coupling coefficients in Fig. 1c associated with the classical angular momentum diagram in Fig. 1a obey the recursion equation[3]

$$(j_1+1)\,a(j_1)\begin{pmatrix} j_1-1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} - (2j_1+1)\,b(j_1)$$

$$\times \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} + j_1 a(j_1+1)\begin{pmatrix} j_1+1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = 0 \tag{5a}$$

where

$$a(j_1) = \{[(j_2+j_3+1)^2 - j_1^2][j_1^2 - (j_2-j_3)^2][j_1^2 - m_1^2]\}^{1/2}, \tag{5b}$$

$$b(j_1) = 2j_1(j_1+1)m_2 + [j_1(j_1+1) + j_2(j_2+1) - j_3(j_3+1)]m_1. \tag{5c}$$

The coupling coefficients in Fig. 2c associated with the classical angular momentum diagram in Fig. 2a are the solution of[3]

$$c(m_2, m_3)\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2-1 & m_3+1 \end{pmatrix} + d(m_2, m_3)\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}$$

$$+ c(m_2+1, m_3-1)\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2+1 & m_3-1 \end{pmatrix} = 0 \tag{6a}$$

where

$$c(m_2, m_3) = [(j_2 - m_2 + 1)(j_2 + m_2)(j_3 + m_3 + 1)(j_3 - m_3)]^{1/2}, \tag{6b}$$

$$d(m_2, m_3) = j_2(j_2+1) + j_3(j_3+1) - j_1(j_1+1) + 2m_2m_3. \tag{6c}$$

In the limit of large angular momentum quantum numbers recursion equations (5) and (6) are algebraically connected with the classical vector diagrams in Figs. 1a and 2a. To demonstrate this asymptotic connection, one may multiply the quantum numbers j_i and m_i by a parameter λ assumed to be large such that all terms in Eqs. (5) and (6) of order $O(\lambda^{-2})$ and smaller may be neglected. Let us apply this approximation to Eq. (5) first. The length of a classical angular momentum vector corresponding to the quantum number λj_i is $\lambda J_i = \lambda j_i + \frac{1}{2}$ (in units $\hbar$). Substituting this in (5b) gives

$$a(j_1) = 4F(\lambda J_1 - \tfrac{1}{2},\ \lambda J_2,\ \lambda J_3)P(\lambda J_1 - \tfrac{1}{2},\ \lambda m_1)$$

where

$$F(a,b,c) = \tfrac{1}{4}[(a+b+c)(-a+b+c)(a-b+c)(a+b-c)]^{1/2}$$

is the area of a triangle $\Delta(a,b,c)$ and $P(r,z) = \sqrt{r^2 - z^2}$ is the x,y component of a vector $\mathbf{r}$ with z component z. $F(\lambda J_1 - \frac{1}{2}, \lambda J_2, \lambda J_3)$ may be factorized in the following way:

$$F(\lambda J_1 - \tfrac{1}{2}, \lambda J_2, \lambda J_3)$$

$$= [F(\lambda J_1 - 1, \lambda J_2, \lambda J_3)F(\lambda J_1, \lambda J_2, \lambda J_3)]^{1/2}$$

$$\times \left[\frac{F(J_1 - (1/2\lambda), J_2, J_3)}{F(J_1 - (1/\lambda), J_2, J_3)}\,\frac{F(J_1 - (1/2\lambda), J_2, J_3)}{F(J_1, J_2, J_3)}\right]^{1/2}$$

$$= [1 + 0(\lambda^{-2})][F(\lambda J_1 - 1, \lambda J_2, \lambda J_3)F(\lambda J_1, \lambda J_2, \lambda J_3)]^{1/2}.$$

Performing this factorization also for $P(\lambda J_1 - \frac{1}{2}, \lambda m_1)$ and $\lambda J_1 + \frac{1}{2}$ gives

$$(j_1+1)\,a(j_1) = [1 + 0(\lambda^{-2})]\,4[\lambda J_1(\lambda J_1 + 1)F(\lambda J_1 - 1, \lambda J_2, \lambda J_3)$$
$$\times F(\lambda J_1, \lambda J_2, \lambda J_3)P(\lambda J_1 - 1, \lambda m_1)P(\lambda J_1, \lambda m_1)]^{1/2}.$$

Substitution of the classical angular momenta λJ_i into (5c) yields

$$b(j_1) = 2\lambda^2 J_1^2 \lambda m_2 + (\lambda^2 J_1^2 + \lambda^2 J_2^2 - \lambda^2 J_3^2)\lambda m_1 - \tfrac{1}{2}\lambda m_2 - \tfrac{1}{4}\lambda m_1$$
$$= [1 + O(\lambda^{-2})][2\lambda^2 J_1^2 \lambda m_2 + (\lambda^2 J_1^2 + \lambda^2 J_2^2 - \lambda^2 J_3^2)\lambda m_1].$$

Neglecting then terms of order $O(\lambda^{-2})$ allows us to write (5a) in the more symmetric form

$$\left(\frac{F(\lambda J_1 - 1, \lambda J_2, \lambda J_3)\,P(\lambda J_1 - 1, \lambda m_1)}{\lambda J_1 - 1}\right)^{1/2} \begin{pmatrix} \lambda j_1 - 1 & \lambda j_2 & \lambda j_3 \\ \lambda m_1 & \lambda m_2 & \lambda m_3 \end{pmatrix}$$

$$- 2\,\frac{2\lambda^2 J_1^2 \lambda m_2 + (\lambda^2 J_1^2 + \lambda^2 J_2^2 - \lambda^2 J_3^2)\lambda m_1}{4F(\lambda J_1, \lambda J_2, \lambda J_3)P(\lambda J_1, \lambda m_1)}$$

$$\times \left(\frac{F(\lambda J_1, \lambda J_2, \lambda J_3)\,P(\lambda J_1, \lambda m_1)}{\lambda J_1}\right)^{1/2} \begin{pmatrix} \lambda j_1 & \lambda j_2 & \lambda j_3 \\ \lambda m_1 & \lambda m_2 & \lambda m_3 \end{pmatrix}$$

$$+ \left(\frac{F(\lambda J_1 + 1, \lambda J_2, \lambda J_3)\,P(\lambda J_1 + 1, \lambda m_1)}{\lambda J_1 + 1}\right)^{1/2}$$

$$\times \begin{pmatrix} \lambda j_1 + 1 & \lambda j_2 & \lambda j_3 \\ \lambda m_1 & \lambda m_2 & \lambda m_3 \end{pmatrix} \approx 0 \tag{7}$$

if one sets in addition $\lambda J_1/[(\lambda J_1 - 1)(\lambda J_1 + 1)]^{1/2} = 1 + O(\lambda^{-2}) \approx 1$.

Keeping the parameter λ explicitly in the following derivations would lead to a somewhat cumbersome notation. We will therefore omit the parameter λ in the remaining formulas and assume instead the quantum numbers $j_i(J_i)$ and m_i to be large.

The classical vector diagram (prism) in Fig. 1a is the geometrical counterpart of the $3j$-coefficients $\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}$. Its triangular base $\Delta(J_1, J_2, J_3)$ contains the sides $j_1 + \frac{1}{2}$, $j_2 + \frac{1}{2}$, and $j_3 + \frac{1}{2}$, the parallel edges L_1, L_2, L_3 are connected with the magnetic quantum numbers through

$$m_1 = L_2 - L_3, \quad m_2 = L_3 - L_1, \quad m_3 = L_1 - L_2. \tag{8}$$

The second base in the x,y plane is the triangle $\Delta[P(j_1, m_1);\ P(j_2, m_2);\ P(j_3, m_3)]$. To emphasize the geometric interpretation of the $3j$-coefficients we define

$$\begin{bmatrix} J_1 & J_2 & J_3 \\ L_1 & L_2 & L_3 \end{bmatrix} \equiv \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}. \tag{9}$$

Since for nonvanishing $3j$-coefficients $m_1 + m_2 = 0$, the three magnetic quantum numbers represent only two independent variables for the $3j$-coefficient. This property is reflected upon the new variables L_1, L_2, and L_3 through

$$\begin{bmatrix} J_1 & J_2 & J_3 \\ L_1 + L & L_2 + L & L_3 + L \end{bmatrix} = \begin{bmatrix} J_1 & J_2 & J_3 \\ L_1 & L_2 & L_3 \end{bmatrix}$$

for any constant L. Addition of such a constant L is geometrically equivalent to a parallel displacement of the x,y plane in Figs. 1a and 2a along the z axis.

The coefficients in Eq. (7) are related algebraically to the classical vector diagram associated with $\begin{bmatrix} J_1 & J_2 & J_3 \\ L_1 & L_2 & L_3 \end{bmatrix}$. The area A of $\Delta[P(j_1, m_1);\ P(j_2, m_2);\ P(j_3, m_3)]$ is

$$A=\frac{F(J_1,J_2,J_3)P(J_1,m_1)}{J_1}\sin\theta_1 \tag{10}$$

where θ_1, the dihedral angle of the prism adjacent to the side J_1, is determined algebraically through[1]

$$\cos\theta_1=[2J_1{}^2m_2+(J_1{}^2+J_2{}^2-J_3{}^2)m_1]/4F(J_1,J_2,J_3)P(J_1,m_1). \tag{11}$$

By virtue of Eqs. (10) and (11), Eq. (7) becomes

$$\left(\frac{A(J_1-1)}{\sin\theta_1(J_1-1)}\right)^{1/2}\begin{bmatrix}J_1-1 & J_2 & J_3\\ L_1 & L_2 & L_3\end{bmatrix}+\left(\frac{A(J_1+1)}{\sin\theta_1(J_1+1)}\right)^{1/2}\times\begin{bmatrix}J_1+1 & J_2 & J_3\\ L_1 & L_2 & L_3\end{bmatrix}$$

$$+2\cos\theta_1(J_1)\left(\frac{A(J_1)}{\sin\theta_1(J_1)}\right)^{1/2}\begin{bmatrix}J_1 & J_2 & J_3\\ L_1 & L_2 & L_3\end{bmatrix}\approx 0. \tag{12}$$

This equation, the asymptotic form of Eq. (5) for large quantum numbers, may be regarded as the semiclassical version of the recursion equation for the $3j$-coefficients in Fig. 1c.

The asymptotic form of recursion equation (6) can also be derived by expansion in terms of the parameter λ introduced above. We will not follow this procedure explicitly, but point out that the approximation taken again neglects only terms of order $O(\lambda^{-2})$. Equations (6b) and (6c) are then approximately

$$c(m_2,m_3)\approx[P(J_2,m_2-1)P(J_2,m_2)P(J_3,m_3)P(J_3,m_3+1)]^{1/2},$$

$$d(m_2,m_3)\approx J_2{}^2+J_3{}^2-J_1{}^2+2m_2m_3,$$

so that (6a) takes the symmetric form

$$[P(J_2,m_2-1)P(J_3,m_3+1)]^{1/2}\begin{pmatrix}j_1 & j_2 & j_3\\ m_1 & m_2-1 & m_3+1\end{pmatrix}$$

$$+[P(J_2,m_2+1)P(J_3,m_3-1)]^{1/2}\begin{pmatrix}j_1 & j_2 & j_3\\ m_1 & m_2+1 & m_3-1\end{pmatrix}$$

$$+\frac{J_2{}^2+J_3{}^2-J_1{}^2+2m_2m_3}{P(J_2,m_2)P(j_3,m_3)}[P(J_2,m_2)P(J_3,m_3)]^{1/2}\times\begin{pmatrix}j_1 & j_2 & j_3\\ m_1 & m_2 & m_3\end{pmatrix}\approx 0. \tag{13}$$

The coefficients in this equation can be related algebraically to the classical vector diagram in Fig. (2a). By means of the identities[1]

$$2A=P(J_2,m_2)P(J_3,m_3)\sin\eta_1 \tag{14}$$

and

$$\cos\eta_1=-\tfrac{1}{2}\frac{J_2{}^2+J_3{}^2-J_1{}^2+2m_2m_3}{P(J_2,m_2)P(J_3,m_3)} \tag{15}$$

where η_1 is the dihedral angle of the angular momentum prism adjacent to the side L_1, Eq. (13) becomes

$$\left(\frac{A(L_1-1)}{\sin\eta_1(L_1-1)}\right)^{1/2}\begin{bmatrix}J_1 & J_2 & J_3\\ L_1-1 & L_2 & L_3\end{bmatrix}+\left(\frac{A(L_1+1)}{\sin\eta_1(L_1+1)}\right)^{1/2}\times\begin{bmatrix}J_1 & J_2 & J_3\\ L_1+1 & L_2 & L_3\end{bmatrix}-2\cos\eta_1(L_1)\left(\frac{A(L_1)}{\sin\eta_1(L_1)}\right)^{1/2}\times\begin{bmatrix}J_1 & J_2 & J_3\\ L_1 & L_2 & L_3\end{bmatrix}\approx 0. \tag{16}$$

Equations (12) and (16) may be formally written as second order difference equations. Defining

$$f(J_1)=(-1)^{j_1+j_2+j_3}\left(\frac{A(J_1)}{\sin\theta_1(J_1)}\right)^{1/2}\begin{bmatrix}J_1 & J_2 & J_3\\ L_1 & L_2 & L_3\end{bmatrix}, \tag{17a}$$

$$g(L_1)=(-1)^{j_1+j_2+j_3}\left(\frac{A(L_1)}{\sin\eta_1(L_1)}\right)^{1/2}\begin{bmatrix}J_1 & J_2 & J_3\\ L_1 & L_2 & L_3\end{bmatrix}, \tag{17b}$$

(12) and (16) are

$$[\Delta^2(J_1)+2-2\cos\theta_1(J_1)]f(J_1)\approx 0, \tag{18a}$$

$$[\Delta^2(L_1)+2-2\cos\eta_1(L_1)]g(L_1)\approx 0, \tag{18b}$$

where the second order difference operator is defined through $\Delta^2(x)h(x)=h(x+1)-2h(x)+h(x-1)$.

The complete *quantum* number domains of Eqs. (18a) and (18b) are confined by the selection rules for $3j$-coefficients. Equation (18a) holds over the J_1-domain

$[j_{1\min}+\frac{1}{2},\ j_{1\max}+\frac{1}{2}]$ with $j_{1\min}=\max\{|m_1|,|j_2-j_3|\}$

and $j_{1\max}=j_2+j_3$.

Equation (18b) holds over the m_2-domain

$[m_{2\min},m_{2\max}]$ with $m_{2\min}=\max\{-j_2,-j_3-m_1\}$

and $m_{2\max}=\min\{j_2,j_3-m_1\}$.

The J_1- and m_2-values occurring in the manifolds of the *classical* vector diagrams in Figs. 1a and 2a, respectively, are further confined to the smaller domains $[J_{1\min},J_{1\max}]$,$[M_{2\min},M_{2\max}]$. In Fig. 1a the smallest and largest J_1-values $J_{1\min}$ and $J_{1\max}$ are assumed in the limit that the triangle $\Delta(J_1,J_2,J_3)$ comes to lie perpendicular to the x,y plane (flat prism!). Hence, $J_{1\min}$ and $J_{1\max}$ are determined as solutions of

$$A^2(J_1)=0 \tag{19}$$

or, alternatively, as solutions of either of the equations

$$\cos\theta(J_1)=\pm 1,\quad \cos\eta_i(J_1)=\pm 1 \tag{20}$$

where θ_i and η_i are the dihedral angles of the prism $\begin{bmatrix}J_1 & J_2 & J_3\\ L_1 & L_2 & L_3\end{bmatrix}$ adjacent to the sides J_i and L_i, respectively. The algebraic expressions for θ_2, θ_3 and η_2, η_3 can be obtained from Eqs. (11) and (15) by circular permutation of the labels of J_1, J_2, J_3 and m_1, m_2, m_3. Clearly, in the limit of a flat prism the dihedral angles are either 0 or π.

In Fig. 2a the smallest and largest m_2-values $M_{2\min}$ and $M_{2\max}$ are also assumed in the limit that $\Delta(J_1,J_2J_3)$ is oriented perpendicular to the x,y plane; i.e., $M_{2\min}$ and $M_{2\max}$ are the solutions of

$$A^2(m_2)=0 \tag{19'}$$

or

$$\cos\theta_i(m_2)=\pm 1,\quad \cos\eta_i(m_2)=\pm 1. \tag{20'}$$

As an illustration we may consider the string of $3j$-coefficients $\begin{pmatrix}j_1 & 100 & 60\\ -10 & 60 & -50\end{pmatrix}$ in Fig. 1, the solution of Eq. (18a). In this case, the full *quantum mechanical domain* is [40. 5, 160. 5], and is divided into the smaller *classical*

domain [47.5, 114.5] and the two complementary *nonclassical domains* [40.5, 46.5] and [115.5, 160.5]. The division into a classical and two nonclassical domains is reflected by the functional behavior of the $3j$-coefficients, as can be seen from Fig. 1c. While progressing along the j_1-domain, the $3j$-coefficients oscillate rapidly in the classical region, but decay monotonically to zero in the outer nonclassical domains. This situation that the domain of Eq. (18a) can be divided into a middle classical region and two outer nonclassical regions applies in general, though in some instances a nonclassical domain may contain only one or no quantum number.

Figure 2b which presents the string of $3j$-coefficients $\begin{pmatrix} 120 & 60 & 70 \\ -10 & m_2 & 10-m_2 \end{pmatrix}$ demonstrates that the partitioning into a classical domain and two distinct nonclassical domains prevails also for the domain of Eq. (18b). For the example chosen, the classical m_2-domain $[-21, 30]$ lies well within the quantum mechanical domain $[-60, 60]$ separating the two nonclassical domains $[-60, -22]$ and $[31, 60]$. Again, the values of the $3j$-coefficients oscillate while progressing through the classical domain and monotonically decay to zero in the nonclassical domains.

How is the existence of classical and nonclassical domains reflected by the difference equations (18a) and (18b)? Over the classical domains the dihedral angles θ_1 and η_1 are real by definition through the geometric formulas (11) and (15). In the nonclassical domains the angles θ_1 and η_1 together with all remaining dihedral angles are complex. This can be verified directly from the expressions (11) and (15), which give absolute values <1 for classically allowed quantum numbers and absolute values >1 for nonclassical quantum numbers. One can also show from Eq. (4) that A^2 is positive in the classical region, zero at the boundaries of the classical domain $J_{1\,\min}$, $J_{1\,\max}$ or $M_{2\,\min}$, $M_{2\,\max}$ as postulated by Eqs. (19) and (19′), and negative in the nonclassical domains. Although the classical vector diagrams do not exist for $3j$-coefficients in the nonclassical domains, the algebraic formulas originating from these diagrams, i.e., Eqs. (4), (10), (11), (14), and (15), remain valid beyond the domain of classically allowed quantum numbers. Of importance for the following, is the fact that the expressions for $\cos\theta_i$ and $\cos\eta_i$ are real in the nonclassical regions, so that the real parts of θ_i and η_i are constant in these regions, and equal to either 0 or π.

To solve Eqs. (18a) and (18b) we start from the observation that the dihedral angles θ_1 and η_1, in the limit of large quantum numbers, vary slowly with J_1 and L_1. To demonstrate this for $\theta_1(J_1)$ one may again employ an expansion in terms of the parameter λ introduced above. From (11)

$$\cos\theta_1(\lambda J_1 \pm 1) = -\frac{2J_1^2 m_2 + [J_1^2 + J_2^2 - J_3^2]m_1 \pm (4J_1 m_2 + 2J_1 m_1)\lambda^{-1} + (2m_2 + m_1)\lambda^{-2}}{4F(J_1, J_2, J_3)\,P(J_1, m_1)} = \cos\theta_1(\lambda J_1) \pm O(\lambda^{-1}) + O(\lambda^{-2}),$$

or

$$\cos\theta_1(\lambda J_1 + 1) - 2\cos\theta_1(\lambda J_1) + \cos\theta_1(\lambda J_1 - 1) = O(\lambda^{-2}). \tag{21}$$

For slowly varying $\theta_1(J_1)$ and $\eta_1(L_1)$ the difference equations (18a) and (18b) can be solved by a discrete analog of the WKB approximation, commonly applied to the solution of the Schrödinger second order differential equation in the semiclassical limit. The "discrete WKB approximation" for *difference* equations is developed in the following section. It should be pointed out here that this approximation applied to Eqs. (18a), (18b) neglects again only terms of order $O(\lambda^{-2})$ and, thus, is consistent with the approximations so far employed.

III. WKB APPROXIMATION APPLIED TO SECOND ORDER DIFFERENCE EQUATIONS

We will now consider the approximate solution of the second order difference equations (18a) and (18b) which may be written

$$[\Delta^2 + 2 - 2\cos k(x)]\,f(x) = 0 \tag{22}$$

where we assume $k(x)$ to be real and $0 \le k(x) \le \pi$. The case of imaginary $k(x)$ can be treated the same way as the case of real $k(x)$ and will be commented upon at the end of this section.

Equation (22) determines $f(x)$ over a discrete set of values x, $x \pm 1$, $x \pm 2$, $\cdots$. However, since x stands for some classical angular momentum variable, Eq. (22) can be assumed to hold over a continuous domain of real numbers x. In this instance the solution of (22) can be viewed as a function over a continuous domain, too, since one may take any x as a starting point for the recursion implied by Eq. (22). As a function of a continuous variable one would expect $f(x)$ to be determined as a solution of a differential equation, rather than a difference equation. Indeed, there exists a second order differential equation which is closely related to the difference equation (22) and which has $f(x)$ as a solution. This holds, however, only within the realm of the semiclassical approximation which had been employed when the difference Eqs. (18a) and (18b) were derived.

We will show now that the differential equation which, within the semiclassical approximation, determines the $f(x)$ in Eq. (22) is

$$\left(\frac{d^2}{dx^2} + k^2(x)\right)\left(\frac{\sin k(x)}{k(x)}\right)^{1/2} f(x) = 0. \tag{23}$$

The semiclassical approximation implies that $k(x)$ is a slowly varying function of x, so that we may neglect in solving Eq. (23) all terms of order $O(k'^2)$ and smaller $[k' = (d/dx)k(x)]$ and also all derivatives of $k(x)$ of order 2 and higher. With $N = [\sin k(x)/k(x)]^{1/2}$ Eq. (23) can then be written

$$f^{(2)} \approx -k^2 f - 2\frac{N'}{N} f^{(1)} \tag{24}$$

since $N''/N \approx 0$ within the approximation stated. We define here $f^{(n)} = (d^n/dx^n)f$. Equations (24) allows us to express all derivatives of $f(x)$ in terms of $f(x)$ itself and its first derivative. It is then readily checked that the third and fourth order derivatives are approximately

$$f^{(3)} \approx -k^2 f^{(1)} - 2kk' f - p k^2 f,$$
$$f^{(4)} \approx k^4 f - 4kk' f^{(1)} - 2p k^2 f^{(1)}$$

where $p = -2N'/N$. The even order derivatives are in general

$$f^{(2n)} \approx (-1)^n [k^{2n} f - 2(n-1) n\, k^{2n-3} k' f^{(1)} - n\, k^{2n-2} p f^{(1)}]. \tag{25}$$

This result permits the evaluation of the difference $\Delta^2 f = f(x+1) - 2f(x) + f(x-1)$ by a Taylor series expansion around x:

$$\Delta^2 f \approx 2 \sum_{n=0}^{\infty} (-1)^n \frac{k^{2n}}{(2n)!} f - 2f - \sum_{n=2}^{\infty} (-1)^n \frac{(2n-2)2n}{(2n)!} k^{2n-3} k' f^{(1)} - \sum_{n=1}^{\infty} (-1)^n \frac{2n}{(2n)!} k^{2n-2} p f^{(1)}$$
$$= [2\cos k(x) - 2] f(x) + \left[\left(\frac{d}{dx} \frac{\sin k(x)}{k(x)}\right) + \frac{\sin k(x)}{k(x)} p(x)\right] \times f^{(1)}(x).$$

The second term on the right-hand side of the last equation vanishes identically since

$$p(x) = -\frac{d}{dx} \ln \frac{\sin k(x)}{k(x)},$$

so that $f(x)$, the solution of the differential Eq. (23), also obeys the difference Equation (22). It may be noted that the derivation carried out here holds also if $k(x)$ is pure imaginary.

The approximation which had been invoked to derive the connection between the difference equation (22) and the differential equation (23) is also employed within the customary WKB approximation applied to these equations. Hence, it must be possible to demonstrate by means of the WKB approximation that (22) and (23) have identical solutions. The application of the WKB approximation to Eq. (23) is standard and we may just state that from it results

$$f(x) = \frac{C}{\sqrt{\sin k(x)}} \cos\left(\int_{x_0}^{x} k(x')\, dx' + \alpha\right) \quad (0 < k < \pi), \tag{26}$$

where α and C are constants to be chosen in accordance with possible boundary and normalization conditions.

The WKB approximation can also be applied to the difference equation (22). One sets

$$f(x) = A(x) \cos[\Omega(x) + \alpha] \tag{27}$$

where the functions $A(x)$ and $\Omega(x)$ to be determined are assumed to be slowly varying such that

$$|\Omega''| \ll 1 \quad \text{and} \quad |A'/A| \ll 1 \tag{28}$$

and all higher derivatives of $\Omega(x)$ and $A(x)$ can be neglected. One has

$$f(x \pm 1) \approx A\left(1 \pm \frac{A'}{A}\right) \cos(\Omega + \alpha \pm \Omega' + \Omega''/2)$$

where

$$\cos\left(\Omega + \alpha \pm \Omega' + \frac{\Omega''}{2}\right) \approx \left[1 - \frac{1}{2}\left(\frac{\Omega''}{2}\right)^2\right] \cos(\Omega + \alpha \pm \Omega') - \frac{\Omega''}{2} \sin(\Omega + \alpha \pm \Omega').$$

Neglecting the terms $(\Omega''/2)^2$ and $(A'/A)(\Omega''/2)$, one obtains

$$f(x \pm 1) \approx A\left(\cos\Omega' \mp \frac{\Omega''}{2} \sin\Omega' \pm \frac{A'}{A} \cos\Omega'\right) \cos(\Omega + \alpha) + A\left(\mp \sin\Omega' - \frac{\Omega''}{2} \cos\Omega' - \frac{A'}{A} \sin\Omega'\right) \sin(\Omega + \alpha),$$

which inserted into Eq. (22) leads to

$$A(2\cos\Omega' - 2\cos k)\cos(\Omega + \alpha), \quad - A\left(\Omega'' \cos\Omega' + 2\frac{A'}{A} \sin\Omega'\right) \sin(\Omega + \alpha) \approx 0,$$

or

$$\Omega' \approx k,$$
$$\frac{A'}{A} \approx -\frac{\Omega''}{2} \frac{\cos\Omega'}{\sin\Omega'}.$$

These last equations determine $\Omega(x)$ and $A(x)$ as

$$\Omega(x) = \int_{x_0}^{x} k(x')\, dx', \tag{29}$$
$$A(x) = C/\sqrt{\sin k(x)}, \tag{30}$$

which together with (27) shows that the WKB solutions of (22) and (23) are indeed identical. This finding holds, of course, only as long as the conditions (28) are met. The condition $|\Omega''| \ll 1$ is equivalent to $|k'| \ll 1$ which had also been assumed in deriving the equivalence of (22) and (23). The condition $|A'/A| \ll 1$ is more restrictive and does not hold for k close to either 0 or π.

The equivalence of Eqs. (22) and (23) within the semi-classical approximation allows a solution which holds uniformly over the entire domain $0 \leq k \leq \pi$. Such solution is obtained by means of the uniform WKB approximation applied to the differential equation (23) setting

$$\left(\frac{\sin k(x)}{k(x)}\right)^{1/2} f(x) = A(x) F(\Omega(x)), \tag{31}$$

where $F(\Omega)$ stands for the regular or the irregular Airy function,[8] the solutions of

$$F''(\Omega) - \Omega F(\Omega) = 0. \tag{32}$$

Inserting (31) into (23) gives by virtue of (32)

$$\left(\frac{A''}{A} + \Omega'^2 \Omega + k^2\right) F(\Omega) + \left(2\frac{A'}{A} \Omega' + \Omega''\right) F'(\Omega) = 0.$$

Assuming now that A' is slowly varying such that $A''/A \approx 0$ leads to the equations

$$\Omega'^2 \Omega + k^2 \approx 0,$$
$$A'/A \approx -\tfrac{1}{2}(\Omega''/\Omega'),$$

the solutions of which are

$$\Omega(x) \approx -\left(\tfrac{3}{2} \left| \int_{x_0}^{x} k(x')\, dx' \right|\right)^{2/3}, \tag{33a}$$
$$A(x) \approx C\{|\Omega(x)|^{1/4}/[k(x)]^{1/2}\} \tag{33b}$$

Finally,

$$f(x) = C \frac{|\Omega|^{1/4}}{[\sin k(x)]^{1/2}} F(\Omega(x)). \tag{34}$$

The integration limit x_0 in (33a) must be either of the zeroes of $\sin k(x)$. In the x region distant from x_0 where $|\Omega(x)|$ is large, i.e., where the solution (26) holds, the expressions (34) and (26) can be identically matched. This follows from the asymptotic behavior of Airy functions[8]: For $|\Omega(x)| \gg 1$ and $\omega = \int_{x_0}^{x} k(x')\,dx'$

$$\sqrt{\pi}\,|\Omega|^{1/4} Ai(-\Omega) = \begin{cases} \cos(\omega - \pi/4) & \text{for } \omega > 0 \\ \cos(\omega + \pi/4) & \text{for } \omega < 0 \end{cases}, \tag{35a}$$

$$\sqrt{\pi}\,|\Omega|^{1/4} Bi(-\Omega) = \begin{cases} \cos(\omega + \pi/4) & \text{for } \omega > 0 \\ \cos(\omega - \pi/4) & \text{for } \omega < 0 \end{cases}. \tag{35b}$$

The derivation above for the equivalence of Eqs. (22) and (23) holds also in the case that $k(x)$ is purely imaginary. The uniform WKB approximation applied to Eq. (23) gives in this case

$$\Omega(x) \approx \left(\tfrac{3}{2} \int_{x_0}^{x} |k(x')|\,dx'\right)^{2/3}, \tag{36a}$$

$$A(x) \approx C\,|\Omega(x)|^{1/4}/|k(x)|^{1/2}, \tag{36b}$$

and hence

$$f(x) = \frac{|\Omega(x)|^{1/4}}{[\sinh k(x)]^{1/2}} F(|\Omega(x)|). \tag{37}$$

IV. SEMICLASSICAL SOLUTION FOR 3j-COEFFICIENTS

The semiclassical solutions for the difference equations (18a) and (18b) have been derived formally, and we will now evaluate explicit expressions for individual 3j-coefficients from these solutions. It is instructive to consider first solution (26) for the difference equations which holds only in the classical domain distant from the classical boundaries $J_{1\,\min}$, $J_{1\,\max}$, and $M_{2\,\min}$, $M_{2\,\max}$. For Eq. (18a) this is

$$\begin{bmatrix} J_1 & J_2 & J_3 \\ L_1 & L_2 & L_3 \end{bmatrix} = (-1)^{j_1+j_2+j_3} \frac{C_1}{\sqrt{A}} \cos[\Omega(J_1) + \alpha] \tag{38}$$

where

$$\Omega(J_1) = \int_{J_{1\,\min}}^{J_1} \theta_1(J_1')\,dJ_1' \tag{39}$$

and for Eq. (18b)

$$\begin{bmatrix} J_1 & J_2 & J_3 \\ L_1 & L_2 & L_3 \end{bmatrix} = (-1)^{j_1+j_2+j_3} \frac{C_2}{\sqrt{A}} \cos[\Phi(L_1) + \beta] \tag{40}$$

where

$$\Phi(L_1) = \int_{L_{1\,\min}}^{L_1} \eta_1(L_1')\,dL_1'. \tag{41}$$

Of course expressions (38) and (40) must be identical and to show this the phase functions $\Omega(J_1)$ and $\Phi(L_1)$ need to be evaluated explicitly. In doing so, we will closely follow Ref. 1.

We had pointed out above that the three variables L_1, L_2, and L_3 in $\begin{bmatrix} J_1 & J_2 & J_3 \\ L_1 & L_2 & L_3 \end{bmatrix}$ are only defined up to a constant which can be added to these variables without changing the value of the 3j-coefficient. Hence, one is free to let L_1, L_2, L_3 go to infinity keeping the differences $m_1 = L_2 - L_3$, $m_2 = L_3 - L_1$, $m_3 = L_1 - L_2$ constant. In this limit the angular momentum prism associated with $\begin{bmatrix} J_1 & J_2 & J_3 \\ L_1 & L_2 & L_3 \end{bmatrix}$ resembles a tetrahedron $T(J_1, J_2, J_3, L_1, L_2, L_3)$ with the triangular base $\Delta(J_1, J_2, J_3)$ and infinite edges L_1, L_2, L_3. In order to evaluate $\Omega(J_1)$ and $\Phi(L_1)$ one can, hence, take advantage of the identity

$$\sum_{i=1}^{3} (J_i\,d\theta_i + L_i\,d\eta_i) = 0, \tag{42}$$

a special case of a theorem first derived by Schlaefli[9] which holds more generally for elliptic tetrahedra. From (42) follows

$$d \sum_{i=1}^{3} (J_i\theta_i + L_i\eta_i) = \sum_{i=1}^{3} \theta_i\,dJ_i + \eta_i\,dL_i$$

or

$$\frac{\partial}{\partial J_k} \sum_{i=1}^{3} (J_i\theta_i + L_i\eta_i) = \theta_k,$$

$$\frac{\partial}{\partial L_k} \sum_{i=1}^{3} (J_i\theta_i + L_i\eta_i) = \eta_k, \tag{43}$$

and one can conclude immediately that $\sum_{i=1}^{3} (J_i\theta_i + L_i\eta_i)$ is the solution of both Eq. (39) and (41).

It should be demonstrated that with this solution (38) and (40) are indeed independent of the absolute lengths of L_1, L_2 and L_3. This follows from the geometrical relationship

$$\eta_1 + \eta_2 + \eta_3 = 2\pi,$$

since then

$$\sum_{i=1}^{3} L_i\eta_i = -\eta_1 m_2 + \eta_2 m_1 + 2\pi L_3.$$

The constant term $2\pi L_3 \pmod{2\pi}$ can be added to the phase constants α and β in (38) and (40), so that we may set

$$\Omega(J_1) = \Phi(L_1) = \sum_{i=1}^{3} J_i\theta_i - \eta_1 m_2 + \eta_2 m_1. \tag{44}$$

(38) and (40) are therefore identical if we set $C_1 = C_2$ and $\alpha = \beta$.

The normalization constant C_1 and the phase constant α still need to be determined. We will derive here only C_1 and must leave the problem of choosing α correctly to a later discussion of the boundary conditions of the 3j-coefficients in the nonclassical domains. The 3j-coefficients represent the matrix elements of the unitary transformation T connecting the basis sets $|j_2 m_2\rangle\,|j_3 m_3\rangle$ and $|(j_2, j_3) j_1 m_1\rangle$. One may exploit the unitary property of T in order to determine the normalization constant C_1:

$$\sum_{J_1} 2J_1 \begin{bmatrix} J_1 & J_2 & J_3 \\ L_1 & L_2 & L_3 \end{bmatrix} \begin{bmatrix} J_1 & J_2 & J_3 \\ L_1' & L_2 & L_3 \end{bmatrix} = \delta_{L_1 L_1'}. \tag{45}$$

In the limit of large quantum numbers this may be written for $L_1 \approx L_1'$ with

$$\begin{bmatrix} J_1 & J_2 & J_3 \\ L_1 & L_2 & L_3 \end{bmatrix} \approx N(J_1, L_1) \cos[\Omega(J_1, L_1) + \alpha],$$

$$\int_{-\infty}^{+\infty} dJ_1\, 2J_1\, N(J_1, L_1)^2 \cos[\Omega(J_1, L_1) + \alpha]$$
$$\times \cos[\Omega(J_1, L_1') + \alpha] \approx \delta(L_1 - L_1').$$

The method of stationary phase can be applied to get

$$\int_{-\infty}^{+\infty} dJ_1\, J_1 N(J_1, L_1)^2 \cos\left((L_1' - L_1)\,\frac{\partial\Omega}{\partial L_1}\right) \approx \delta(L_1 - L_1')$$

which holds only if

$$N = \pm\left(\frac{1}{\pi J_1}\left|\frac{\partial^2\Omega}{\partial J_1 \partial L_1}\right|\right)^{1/2}. \tag{46}$$

To evaluate $\partial^2\Omega/\partial J_1\partial L_1 = \partial\eta_1/\partial J_1$ we note that

$$\cos\eta_1 = -4\,\frac{\partial A^2/\partial J_1^2}{P(J_2, m_2)P(J_3, m_3)}$$

which can be derived from Eq. (4) and Eq. (15). Together with (14) this gives

$$\frac{\partial\eta_1}{\partial J_1} = -\frac{J_1}{2A} \tag{47}$$

and, hence,

$$N = \pm\left(\frac{1}{2\pi A}\right)^{1/2}.$$

This result is in agreement with (38) and (40) for $C_1 = 1/\sqrt{2\pi}$. Evidently, this determination of the normalization constant also yields the correct functional dependence $1/\sqrt{A}$ for N. It is well known that the phase function of a semiclassical operator alone determines the normalization factor N. Our result supports therefore (38) as the correct semiclassical $3j$-coefficient. We have finally for the $3j$-coefficients in the classical domain

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = (-1)^{j_1+j_2+j_3}\,\frac{1}{\sqrt{2\pi A}}$$
$$\times\cos(J_1\theta_1 + J_2\theta_2 + J_3\theta_3 - \eta_1 m_2 + \eta_2 m_1 + \alpha). \tag{48}$$

This result, except for the phase constant α to be determined yet, is identical with the expression of Ponzano and Regge and that derived by Miller from the correspondence relationships of classical and quantum mechanics.

The expression for semiclassical $3j$-coefficients uniformly valid over the entire quantum number domain can be evaluated from the solutions (34) and (37) of the difference equation (22). These solutions are valid even near the classical turning points $J_{1\,\min}$ and $J_{1\,\max}$ ($M_{2\,\min}$ and $M_{2\,\max}$) provided that $k(J_{1\,\min})$, etc. vanish. However, in Eqs. (18a) and (18b) θ_1 and η_1 are either 0 or π at the turning points. In the latter case, a phase transformation has to be employed which induces the proper turning point behavior of the difference equation.

For the following, it may be sufficient to consider Eq. (18a) only, since (18b) would lead to an identical solution. We define the phase function

$$\Omega_0 = J_1\theta_1^0 + J_2\theta_2^0 + J_3\theta_3^0 + m_1\eta_2^0 - m_2\eta_1^0 \tag{49}$$

where

$$\theta_i^0 = \begin{cases} 0 & \text{if } 0 \leq \mathrm{Re}\theta_i \leq \pi/2 \\ \pi & \text{if } \pi/2 < \mathrm{Re}\theta_i \leq \pi \end{cases}, \tag{50a}$$

$$\eta_i^0 = \begin{cases} 0 & \text{if } 0 \leq \mathrm{Re}\eta_i \leq \pi/2 \\ \pi & \text{if } \pi/2 < \mathrm{Re}\eta_i \leq \pi \end{cases}. \tag{50b}$$

$\mathrm{Re}\theta_i$ ($\mathrm{Re}\eta_i$) stands here for the real part of $\theta_i(\eta_i)$. For all quantum mechanically allowed J_i and m_i, Ω_0 thus defined is either an integer or a half-integer multiple of π. A simple criterion can be found which distinguishes between these two possibilities if $J_1 \approx J_{1\,\min}$ or $J_1 \approx J_{1\,\max}$:

$$\Omega_0/\pi \text{ half integer (integer)} \iff \Omega - \Omega_0 \text{ positive (negative)}. \tag{51}$$

To derive this rule one may express $\Omega - \Omega_0$ through $\int(\theta_1 - \theta_1^0)dJ_1$. For $J_1 \approx J_{1\,\min}$

$$\Omega - \Omega_0 = \int_{J_{1\,\min}}^{J_1} (\theta_1 - \theta_1^0)\, dJ_1$$

and since θ_1 is a monotonous function in the neighborhood of $J_{1\,\min}$ it follows from $\theta_1(J_{1\,\min}) = \theta_1^0$

$$\Omega - \Omega_0 = \begin{cases} >0 & \text{if } \theta_1^0 = 0 \\ <0 & \text{if } \theta_1^0 = \pi \end{cases}.$$

However, for $\theta_1^0(J_{1\,\min}) = 0$

$$\begin{bmatrix} \theta_1^0 & \theta_2^0 & \theta_3^0 \\ \eta_1^0 & \eta_2^0 & \eta_3^0 \end{bmatrix} = \begin{bmatrix} 0 & \pi & 0 \\ \pi & 0 & \pi \end{bmatrix} \tag{52a}$$

and Ω_0/π is half integer, and for $\theta_1^0(J_{1\,\min}) = \pi$

$$\begin{bmatrix} \theta_1^0 & \theta_2^0 & \theta_3^0 \\ \eta_1^0 & \eta_2^0 & \eta_3^0 \end{bmatrix} = \begin{bmatrix} \pi & 0 & \pi \\ \pi & 0 & \pi \end{bmatrix} \tag{52b}$$

and Ω_0/π is integer. Similarly, for $J_1 \approx J_{1\,\max}$

$$\Omega - \Omega_0 = -\int_{J_1}^{J_{1\,\max}} (\theta_1 - \theta_1^0)\, dJ_1$$

is negative for $\theta_1^0 = 0$ and positive for $\theta_1^0 = \pi$. But in this case one finds for $\theta_1^0(J_{1\,\max}) = 0$

$$\begin{bmatrix} \theta_1^0 & \theta_2^0 & \theta_3^0 \\ \eta_1^0 & \eta_2^0 & \eta_3^0 \end{bmatrix} = \begin{bmatrix} 0 & \pi & \pi \\ 0 & \pi & \pi \end{bmatrix}, \tag{52c}$$

i.e., Ω_0/π integer, and for $\theta_1^0(J_{1\,\max}) = \pi$

$$\begin{bmatrix} \theta_1^0 & \theta_2^0 & \theta_3^0 \\ \eta_1^0 & \eta_2^0 & \eta_3^0 \end{bmatrix} = \begin{bmatrix} \pi & 0 & 0 \\ 0 & \pi & \pi \end{bmatrix}, \tag{52d}$$

i.e., Ω_0/π half integer. Equations (52a)–(52d) are readily checked geometrically by drawing the corresponding angular momentum diagrams.

At the boundaries of the classical domain $J_{1\,\min}$, $J_{1\,\max}$ the dihedral angles θ_i and η_i are either 0 or π. Since the dihedral angles are slowly varying functions of the quantum numbers, the angles θ_i^0 and η_i^0 are constant near $J_{1\,\min}$ and $J_{1\,\max}$. One may then set for the solution f of Eq. (18a)

$$f = \cos\Omega_0 f_1 - \sin\Omega_0 f_2 \tag{53}$$

and show that in the regions of constant $\theta_1^0 = 0$ or π the functions f_1 and f_2 must satisfy the modified difference equation.

$$[\Delta^2(J_1) + 2 - 2\cos(\theta_1 - \theta_1^0)]\, f_i(J_1) \approx 0. \tag{54}$$

The new representation (53) for the $3j$-coefficients together with this new difference equation, has favorable behavior at the boundaries of the classical region, and will also facilitate the evaluation of the $3j$-coefficients in the nonclassical region.

In the classical domain distant from $J_{1\,\min}$ and $J_{1\,\max}$ $f(J_1)$ may be written according to Eqs. (17a) and (48)

$$f=[2\pi\sin\theta_1]^{-1/2}[\cos\Omega_0\cos(\Omega-\Omega_0+\alpha) - \sin\Omega_0\sin(\Omega-\Omega_0+\alpha)] \tag{55}$$

and f_1 and f_2 must be chosen such that (53) identically matches this expression for $J_{1\,\min}\ll J_1\ll J_{1\,\max}$.

It had been shown in Sec. 3 that the solutions of (54) in the semiclassical limit must obey the differential equation

$$\left(\frac{d^2}{dJ_1^2}-(\theta_1-\theta_1^0)^2\right)\left(\frac{\sin(\theta_1-\theta_1^0)}{\theta_1-\theta_1^0}\right)^{1/2} f_i(J_1)=0 \tag{56}$$

where J_1 is now assumed to be a continuous variable. In this equation $(\theta_1-\theta_1^0)^2$ is positive in the classical domain, zero at $J_1=J_{1\,\min}$, $J_{1\,\max}$ and negative in the nonclassical domain. According to Eqs. (33a,b) and (34) the solutions of (56) are in the classical domain

$$\frac{Z^{1/4}}{(\sin\theta_1)^{1/2}}\,\mathrm{Ai}(-Z) \quad\text{and}\quad \frac{Z^{1/4}}{(\sin\theta_1)^{1/2}}\,\mathrm{Bi}(-Z)$$

where $Z=(\frac{3}{2}|\Omega-\Omega_0|)^{2/3}$. The solutions in the nonclassical domain obtained by analytical continuation are

$$\frac{Z^{1/4}}{(\sinh|\theta_1-\theta_1^0|)^{1/2}}\,\mathrm{Ai}(Z) \quad\text{and}\quad \frac{Z^{1/4}}{(\sinh|\theta_1-\theta_1^0|)^{1/2}}\,\mathrm{Bi}(Z).$$

To identify the functions f_1 and f_2 the boundary conditions to be imposed on the $3j$-coefficients have to be taken into consideration. Evidently, the $3j$-coefficients must decay monotonically to zero in the nonclassical domain so that the irregular Airy function $\mathrm{Bi}(J)$ which exhibit an exponential increase must be rejected in this domain. This boundary condition is satisfied if we set for $J_{1\,\min}\le J_1\le J_{1\,\max}$

$$f=\frac{Z^{1/4}}{(\sin\theta_1)^{1/2}} \times\begin{cases} a\cos\Omega_0\,\mathrm{Ai}(-Z)-b\sin\Omega_0\,\mathrm{Bi}(-Z), & \Omega-\Omega_0<0\\ a'\cos\Omega_0\,\mathrm{Bi}(-Z)-b'\sin\Omega_0\,\mathrm{Ai}(-Z), & \Omega-\Omega_0>0\end{cases} \tag{57}$$

which analytically continued into the nonclassical domain is

$$f=\frac{Z^{1/4}}{(\sinh|\theta_1-\theta_1^0|)^{1/2}} \times\begin{cases} a\cos\Omega_0\,\mathrm{Ai}(Z)-b\sin\Omega_0\,\mathrm{Bi}(Z), & \Omega-\Omega_0<0\\ a'\cos\Omega_0\,\mathrm{Bi}(Z)-b'\sin\Omega_0\,\mathrm{Ai}(Z), & \Omega-\Omega_0>0.\end{cases} \tag{58}$$

Over the discrete set of quantum mechanically allowed J_i and m_i (i.e., J_i, m_i either integer or half integer) the factors $\sin\Omega_0$ and $\cos\Omega_0$ multiplying the irregular Airy functions $\mathrm{Bi}(Z)$ vanish in the nonclassical domains as had been postulated. (57) becomes for $J_{1\,\min}\ll J_1\ll J_{1\,\max}$

$$f=\frac{1}{(\pi\sin\theta_1)^{1/2}} \times\begin{cases} a\cos\Omega_0\cos(\Omega-\Omega_0+\pi/4)-b\sin\Omega_0\sin(\Omega-\Omega_0+\pi/4), & \Omega-\Omega_0<0\\ a'\cos\Omega_0\cos(\Omega-\Omega_0+\pi/4)-b'\sin\Omega_0\sin(\Omega-\Omega_0+\pi/4), & \Omega-\Omega_0>0.\end{cases} \tag{59}$$

The condition that this expression must identically match (55) is met for $\alpha=\pi/4$ and $a=b=a'=b'=S/\sqrt{2}$ where $S=\pm1$, is an arbitrary phase factor. The $3j$-coefficients are then finally

$$\begin{pmatrix} j_1 & j_2 & j_3\\ m_1 & m_2 & m_3\end{pmatrix} =S(-1)^{j_1+j_2+j_3}\,\frac{Z^{1/4}}{\sqrt{2A}} \times\begin{cases} \cos\Omega_0\,\mathrm{Ai}(-Z)-\sin\Omega_0\,\mathrm{Bi}(-Z), & \Omega-\Omega_0<0\\ \cos\Omega_0\,\mathrm{Bi}(-Z)-\sin\Omega_0\,\mathrm{Ai}(-Z), & \Omega-\Omega_0>0\end{cases} \tag{60}$$

in the classical domain and

$$\begin{pmatrix} j_1 & j_2 & j_3\\ m_1 & m_2 & m_3\end{pmatrix} =S(-1)^{j_1+j_2+j_3}\,\frac{Z^{1/4}}{\sqrt{2|A|}} \times\begin{cases} \cos\Omega_0\,\mathrm{Ai}(Z)-\sin\Omega_0\,\mathrm{Bi}(Z), & \Omega-\Omega_0<0\\ \cos\Omega_0\,\mathrm{Bi}(Z)-\sin\Omega_0\,\mathrm{Ai}(Z), & \Omega-\Omega_0>0\end{cases} \tag{61}$$

in the nonclassical domain. The phase factor S may be defined according to the phase convention of Wigner

$$\mathrm{sgn}\left\{\begin{pmatrix} j_2+j_3 & j_2 & j_3\\ m_1 & m_2 & m_3\end{pmatrix}\right\}=(-1)^{j_2-j_3-m_1}. \tag{62}$$

For $J_1=j_2+j_3+\frac{1}{2}$ in the nonclassical domain $\mathrm{Ai}(Z)>0$ and

$$\cos\Omega_0(\Omega-\Omega_0>0)=-\sin\Omega_0(\Omega-\Omega_0<0) =(-1)^{j_2+j_3+m_1+1},$$

so that

$$\mathrm{sgn}\left\{\begin{pmatrix} j_1 & j_2 & j_3\\ m_1 & m_2 & m_3\end{pmatrix}\right\}=S(-1)^{j_2+j_3+m_1+1}$$

and $S=(-1)^{-2j_3+1}$.

The $3j$-coefficients (60) and (61) had been derived as the semiclassical solutions of the recursion equations (5) and (6). These recursion equations can be derived from the following eigenvalue problems:

(5) : diagonalization of J_{2z} in the basis $|(j_2,j_3)j_1m_1\rangle$: eigenvalues $m_2=m_{2\,\min}+n$, $n=0,1,2,\cdots$,

(6) : diagonalization of $(\mathbf{J}_2+\mathbf{J}_3)^2$ in the basis $|j_2m_2\rangle\,|j_3m_3\rangle$: eigenvalues $j_1(j_1+1)$ where $j_1=j_{1\,\min}+n$, $n=0,1,2,\cdots$.

The eigenvectors i.e., the rows and columns of the unitary transformation matrix T defined in (1) and (2), are then determined except for overall constant factors by the recursion equations (5) and (6). In solving these recursion equations one may treat the eigenvalues m_2 and $j_1(j_1+1)$ as unknowns to be determined through the

boundary conditions to be imposed on the solutions (60), (61). The boundary conditions that the $3j$-coefficients decay to zero in the nonclassical domains can only be satisfied if $J_i(j_i)$ and m_i are either integer or half-integer for otherwise the coefficients $\sin\Omega_0(\Omega-\Omega_0<0)$ and $\cos\Omega_0(\Omega-\Omega_0>0)$ in (61) multiplying the exponentially increasing irregular Airy functions $\mathrm{Bi}(Z)$ do not vanish. The semiclassical $3j$-coefficients exhibit, hence, a typical quantum character in the variables J_1 and $m_2(m_3)$. Since the semiclassical solution (60), (61) is independent of the recursion equation from which the derivation is started, this quantization property must hold for all the variables of the $3j$-coefficients. Furthermore, it is important to notice that the set of semiclassical quantum numbers J_i and m_i being thus defined as the allowed values J_i and m_i coincides with the set of quantum mechanical quantum numbers $j_i+\frac{1}{2}$, m_i.

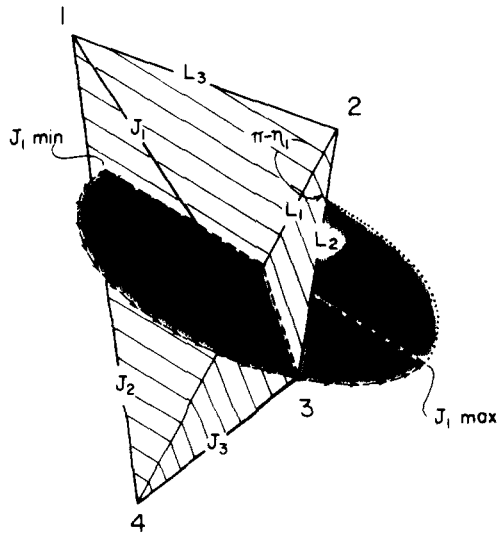
V. SEMICLASSICAL $6j$-COEFFICIENTS

The $6j$-coefficients $\{{}^{j_1\ j_2\ j_3}_{l_1\ l_2\ l_3}\}$ define a unitary transformation T

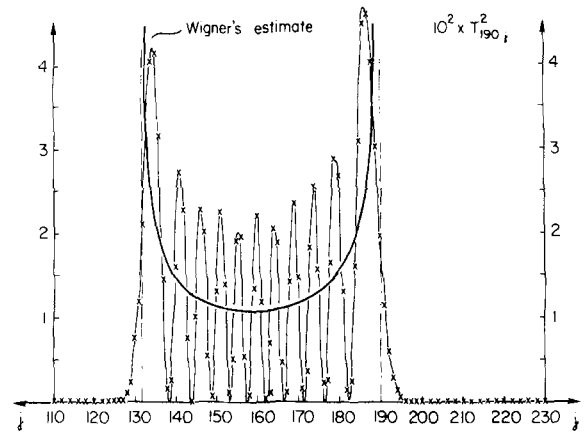
$$T_{j_1 l_1}=[(2j_1+1)(2l_1+1)]^{1/2}\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix} \tag{63}$$

between two total angular momentum states with the common internal angular momenta l_2, l_3, j_3 coupled in different ways to the same total angular momentum j_2.[7] One state is coupled according to the scheme $(j_3, (l_2, l_3)j_1; j_2)$, i.e., l_2 and l_3 are coupled to the *intermediate* j_1 which in turn is coupled with j_3 to give j_2. The other state is coupled according to the scheme $(l_3(l_2, j_3)l_1; j_2)$, i.e., l_2, j_3 are coupled to the *intermediate* l_1 which in turn is coupled with l_3 to give j_2. Classically this corresponds to two ways of addition of angular momentum vectors, namely $\mathbf{L}_2+\mathbf{L}_3=\mathbf{J}_1, \mathbf{J}_1+\mathbf{J}_3=\mathbf{J}_2$ and $\mathbf{L}_2+\mathbf{J}_3=\mathbf{L}_1$, $\mathbf{L}_1+\mathbf{L}_3=\mathbf{J}_2$. This angular momentum coupling situation is illustrated by the classical vector diagram in figure 3a where the classical angular momentum vectors are defined in the usual way, $J_i=j_i+\frac{1}{2}$ and $L_i=l_i+\frac{1}{2}$.

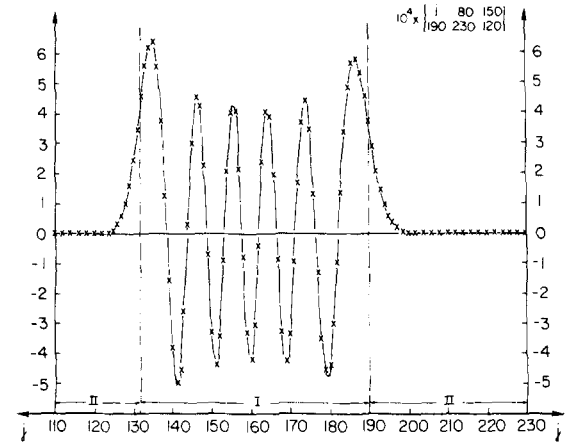
Let us now assume as fixed the lengths of the classical angular momenta $\mathbf{J}_2, \mathbf{J}_3$, $\mathbf{L}_1, \mathbf{L}_2$ and $\mathbf{L}_3$ in Fig. 3a and, accordingly, the quantum numbers j_2, j_3, l_1, l_2, l_3 in $\{{}^{j_1\ j_2\ j_3}_{l_1\ l_2\ l_3}\}$ and let us consider the $6j$-coefficients for all possible j_1 together with the corresponding classical vector diagrams. The manifold of all possible classical vector diagrams is generated by moving vertex 3 of the angular momentum vector tetrahedron in Fig. 3a along the circle indicated whereby the angle η_1 goes from 0 to π. J_1 assumes then all values which correspond to the varying distances between vertex 1 and 3. To complete the correspondence between $6j$-coefficients and classical vector diagrams it may be recalled that the matrix elements defined in (63) squared $T_{j_1 l_1}{}^2$ are to be interpreted as the quantum mechanical probability for l_2, l_3 to be coupled to give j_1, or, conversely, that $J_1-\frac{1}{2} \leq |\mathbf{L}_2+\mathbf{L}_3| \leq J_1+\frac{1}{2}$. On the basis of this interpretation Wigner[5] established an approximate functional relationship between $6j$-coefficients and the associated classical angular momentum tetrahedra. From the assumption that each angle η_1 in Fig. 3a is equally likely, he estimated for the probability $T_{j_1 l_1}{}^2$



(a)



(b)



(c)

FIG. 3. The series of $6j$-coefficients $\{{}^{j_1\ j_2\ j_3}_{l_1\ l_2\ l_3}\}$, $j_1\min \leq j_1 \leq j_1\max$, in (c) and the corresponding manifold of classical angular momentum tetrahedra ($J_1=j_1+\frac{1}{2}$, $L_i=l_i+\frac{1}{2}$) generated by rotation of vertex 3 around the shaded circle. The quantum mechanical probability distribution $(2j_1+1)(2l_1+1)\{{}^{j_1\ j_2\ j_3}_{l_1\ l_2\ l_3}\}^2$ for the occurrence of the classical vector diagram in (a) are compared with Wigner's semiclassical estimate $(2j_1+1)(2l_1+1)/24\pi V$.

$$T_{j_1 l_1}{}^2 = (2j_1+1)(2l_1+1)/24\pi V \tag{64}$$

where V is the volume of the angular momentum tetrahedron given by the Cayley determinant

$$V^2(J_1) = \frac{1}{288}\begin{vmatrix} 0 & L_1^2 & L_2^2 & L_3^2 & 1 \\ L_1^2 & 0 & J_3^2 & J_2^2 & 1 \\ L_2^2 & J_3^2 & 0 & J_1^2 & 1 \\ L_3^2 & J_2^2 & J_1^2 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}. \tag{65}$$

Equation (64) provides only an estimate for the absolute magnitudes of the $6j$-coefficients. Figure 3c which presents the series of $6j$-coefficients $\left\{\begin{smallmatrix} j_1 & 80 & 150 \\ 190 & 236 & 120 \end{smallmatrix}\right\}$ illustrates that the $6j$-coefficients oscillate rapidly while j_1 is progressing along the circle of classically allowed j_1-values in Fig. 3a. Figure 3b which compares the corresponding probabilities $T_{j_1 l_1}{}^2$ with Wigner's estimate (64) demonstrates that (*it*) holds only as an average over several neighboring quantum numbers. The exact progression of the $6j$-coefficients is, however, determined by the recursion equation[3]

$$j_1 g(j_1+1)\begin{Bmatrix} j_1+1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix} + h(j_1)\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix} + (j_1+1)\, g(j_1)\begin{Bmatrix} j_1-1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix} = 0 \tag{66a}$$

where

$$g(j_1) = \{[(j_2+j_3+1)^2 - j_1^2][j_1^2 - (j_2-j_3)^2][(l_2+l_3+1)^2 - j_1^2] \times [j_1^2 - (l_2-l_3)^2]\}^{1/2}, \tag{66b}$$

$$h(j_1) = (2j_1+1)\{j_1(j_1+1)[-j_1(j_1+1)+j_2(j_2+1)+j_3(j_3+1)] + l_2(l_2+1)[j_1(j_1+1)+j_2(j_2+1)-j_3(j_3+1)] + l_3(l_3+1)[j_1(j_1+1)-j_2(j_2+1)+j_3(j_3+1)] - 2j_1(j_1+1)\, l_1(l_1+1)\}. \tag{66c}$$

We will show in the remainder of this section that for large angular momentum quantum numbers this recursion equation allows us to refine Wigner's result (64).

In the limit of large quantum numbers recursion equation (66) takes on a more symmetric form which allows a geometrical interpretation in terms of the classical vector diagram 3a. This asymptotic recursion equation is again derived through a dilation of the angular momentum quantum numbers (i.e., $j_i, l_i \to \lambda j_i, \lambda l_i$) and a consecutive expansion of (66) in terms of λ neglecting terms of order $O(\lambda^{-2})$ and smaller, assuming that λ is large. Thus (66b) and (66c) may be written

$$(\lambda j_1+1)\, g(\lambda j_1) = [1+O(\lambda^{-2})]\, 16[\lambda J_1(\lambda J_1+1)\, F(\lambda J_1, \lambda J_2, \lambda J_3) \times F(\lambda J_1, \lambda L_2, \lambda L_3)\, F(\lambda J_1 - 1, \lambda J_2, \lambda J_3) \times F(\lambda J_1 - 1, \lambda L_2, \lambda L_3)]^{1/2}$$

and

$$h(\lambda j_1) = -2\lambda J_1[1+O(\lambda^{-2})][2\lambda^2 J_1^2 \lambda^2 L_1^2 - \lambda^2 J_1^2(-\lambda^2 J_1^2 + \lambda^2 L_2^2 + \lambda^2 L_3^2) - \lambda^2 J_2^2(\lambda^2 J_1^2 + \lambda^2 L_2^2 - \lambda^2 L_3^2) - \lambda^2 J_3^2(\lambda^2 J_1^2 - \lambda^2 L_2^2 + \lambda^2 L_3^2)].$$

Neglecting terms of order $O(\lambda^{-2})$ allows us then to bring (66a) into the following symmetric form:

$$\left(\frac{F(\lambda J_1-1, \lambda J_2, \lambda J_3)\, F(\lambda J_1 - 1,\ \lambda L_2, \lambda L_3)}{\lambda J_1 - 1}\right)^{1/2} \begin{Bmatrix} \lambda j_1 - 1 & \lambda j_2 & \lambda j_3 \\ \lambda l_1 & \lambda l_2 & \lambda l_3 \end{Bmatrix} + \left(\frac{F(\lambda J_1+1, \lambda J_2, \lambda J_3)\, F(\lambda J_1+1, \lambda L_2, \lambda L_3)}{\lambda J_1 + 1}\right)^{1/2} \begin{Bmatrix} \lambda j_1 + 1 & \lambda j_2 & \lambda j_3 \\ \lambda l_1 & \lambda l_2 & \lambda l_3 \end{Bmatrix} - 2\,\frac{2J_1^2L_1^2 - J_1^2(-J_1^2+L_2^2+L_3^2) - J_2^2(J_1^2+L_2^2-L_3^2) - J_3^2(J_1^2-L_2^2+L_3^2)}{16\, F(J_1, J_2, J_3)\, F(J_1, L_2, L_3)} \times \left(\frac{F(\lambda J_1, \lambda J_2, \lambda J_3)\, F(\lambda J_1, \lambda L_2, \lambda L_3)}{\lambda J_1}\right)^{1/2} \begin{Bmatrix} \lambda j_1 & \lambda j_2 & \lambda j_3 \\ \lambda l_1 & \lambda l_2 & \lambda l_3 \end{Bmatrix} \approx 0. \tag{67}$$

We will omit the parameter λ in the following derivation to avoid a somewhat cumbersome notation. Doing so, we assume the quantum numbers j_i and l_i to be large. It is again pointed out that all the following approximations taken to solve Eq. (67) are consistent with the neglect of terms $O(\lambda^{-2})$.

The coefficients in Eq. (67) are related algebraically to the classical vector diagrams in Fig. 3(a). This can be recognized by virtue of the relations

$$\tfrac{3}{2} V J_1 = F(J_1, J_2, J_3)\, F(J_1, L_2, L_3) \sin\theta_1 \tag{68}$$

and[1]

$$\cos\theta_1 = \frac{2J_1^2L_1^2 - J_1^2(-J_1^2+L_2^2+L_3^2) - J_2^2(J_1^2+L_2^2-L_3^2) - J_3^2(J_1^2-L_2^2+L_3^2)}{16\, F(J_1, J_2, J_3)\, F(J_1, L_2, L_3)}. \tag{69}$$

(67) may then be formally written as a second order difference equation

$$[\Delta^2(J_1)+2-2\cos\theta_1]\,f(j_1)\approx 0 \tag{70}$$

where

$$f(J_1)=\left(\frac{V}{\sin\theta_1}\right)^{1/2}\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix}. \tag{71}$$

This difference equation is valid over the domain $[j_{1\,\min}, j_{1\,\max}]$ of the j_1-manifold of $6j$-coefficients $\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix}$. $j_{1\,\min}$ and $j_{1\,\max}$ are determined as the smallest and largest j_1-values satisfying the triangular conditions for $\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix}$, i.e., $j_{1\,\min}=\max\{|j_2-j_3|,\ |l_2-l_3|\}$ and $j_{1\,\max}=\min\{j_2+j_3, l_2+l_3\}$. Obviously, the J_1-values occuring in the manifold of the classical vector diagrams in Fig. 3a also lie within a finite interval $[J_{1\,\min}, J_{1\,\max}]$. The smallest and largest classically allowed J_1-values, $J_{1\,\min}$ and $J_{1\,\max}$, are those for which all angular momentum vectors in Fig. 3a happen to lie within a single plane (flat tetrahedron). Hence, $J_{1\,\min}$ and $J_{1\,\max}$ are determined as solutions of either of the equations

$$\cos\theta_i(J_1)=\pm 1,\quad \cos\eta_i(J_1)=\pm 1 \tag{72a}$$

where θ_i and η_i are the dihedral angles of the angular momentum tetrahedron adjacent to the edges J_i and L_i, respectively. The algebraic expressions for $\theta_2, \theta_3, \eta_1, \eta_2, \eta_3$ can be obtained from (69) by permutation of the labels of J_i and L_i and by permutation of J and L. Clearly, in the limit of a flat tetrahedron the dihedral angles are all either 0 or π. Alternatively, $J_{1\,\min}$ and $J_{1\,\max}$ may be determined as solutions of

$$V^2(J_1)=0. \tag{72b}$$

To compare the quantum mechanical domain $[j_{1\,\min}, j_{1\,\max}]$ and the classical domain $[J_{1\,\min}, J_{1\,\max}]$ we turn to Fig. 3c which presents the whole string of $6j$-coefficients $\begin{Bmatrix} j_1 & 80 & 150 \\ 190 & 230 & 120 \end{Bmatrix}$. The classical j_1-domain is for this case [131,190] and lies well within the domain [110,230] of all quantum mechanically allowed j_1-values. Hence, the quantum mechanical domain [110,230] is divided into a classical domain [131, 190] and two nonclassical domains [110, 130] and [191, 230]. This situation that the quantum mechanical j_1-domain of a series of $6j$-coefficients can be divided into a middle classical domain and two outer nonclassical domains applies in general, though in some instances the nonclassical domains may contain only a few or no j_1-quantum numbers. The division into classical and nonclassical domains has an important meaning which is reflected by the functional behavior of the $6j$-coefficients, as can be seen from Fig. 3c. The $6j$-coefficients as one progresses along the j_1-domain oscillate rapidly in the classical domain and decay to zero in the nonclassical domains.

The existence of classical and nonclassical domains is also reflected by the difference equation (70). Over the classical domain $[J_{1\,\min}, J_{1\,\max}]$ the dihedral angle $\theta_1(J_1)$ is real by virtue of its geometrical meaning, but in the nonclassical domains the angle $\theta_1(J_1)$ together with the remaining dihedral angles is complex. This can be verified directly from the algebraic expression for $\cos\theta_1$ given in Eq. (69) which gives values between -1 and $+1$ for classical J_1 and values exceeding -1 and $+1$ for nonclassical J_1. Equation (65) reveals that $V^2(J_1)$ is positive in the classical domain, zero at the classical limits $J_{1\,\min}$ and $J_{1\,\max}$, and negative in the nonclassical domains. Hence, to a $6j$-coefficient in the non-classical domain corresponds an angular momentum vector diagram with complex dihedral angles and imaginary volume.

To solve the difference equation (70) we start as before in the case of the $3j$-coefficients from the observation that for large quantum numbers $\theta_1(J_1)$ is a slowly varying function of J_1, since again

$$\cos\theta_1(\lambda J_1+1,\ \lambda J_2,\ \cdots)=\cos\theta_1(J_1+1/\lambda,\ J_2,\ \cdots),$$

as can be readily checked from Eq. (69). Hence, the "discrete" WKB approximation derived in Sec. 3 may be employed to solve Eq. (70).

In the classical domain $J_{1\,\min}\ll J_1\ll J_{1\,\max}$ the solution as given by (26) is

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix}=\frac{C}{\sqrt{V}}\cos[\Omega(J_1)+\alpha] \tag{73}$$

where

$$\Omega(J_1)=\int_{J_{1\,\min}}^{J_1}\theta_1(J_1')\,dJ_1'. \tag{74}$$

$\Omega(J_1)$ is again readily evaluated by means of Eq. (43) which holds also for the angular momentum tetrahedron in Fig. 3a:

$$\Omega(J_1)=\sum_{i=1}^{3}(J_i\theta_i+L_i\eta_i). \tag{75}$$

To determine the normalization constant C in (73) the unitary property of $6j$-coefficients

$$\sum_{j_1} T_{j_1 l_1}T_{j_1 l_1'}=\delta_{l_1 l_1'} \tag{76}$$

can be employed. With

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix}=N(J_1,L_1)\cos[\Omega(J_1,L_1)+\alpha] \tag{77}$$

for $L_1\approx L_1'$, we have for large quantum numbers

$$\int dJ_1\,4J_1L_1\,N^2(J_1,L_1)\cos[\Omega(J_1,L_1)+\alpha]$$

$$\times\cos[\Omega(J_1,L_1')+\alpha]\approx\delta(L_1-L_1').$$

The method of stationary phase applied to this integral gives

$$\int dJ_1\,2J_1L_1\,N^2(J_1,L_1)\cos\left((L_1'-L_1)\frac{\partial\Omega}{\partial L_1}\right)\approx\delta(L_1-L_1')$$

which holds only if

$$N=\left(\frac{1}{2J_1L_1\pi}\left|\frac{\partial\Omega}{\partial J_1\partial L_1}\right|\right)^{1/2}.$$

However, by virtue of Eq. (43) and $\partial\eta_1/\partial J_1=J_1L_1/6V$ this is

$$N=1/\sqrt{12\pi V}\ , \tag{78}$$

so that $C=1/\sqrt{12\pi}$ is the correct normalization constant.

The validity of solution (73) is subject to the condition $N^{-1}|\partial N/\partial J_1|\ll 1$. This condition cannot be satisfied for small V in the neighborhood of $J_{1\,\min}$ and $J_{1\,\max}$. An expression for the semiclassical $6j$-coefficients which is

also valid in these regions is found by performing first a suitable phase transformation of the solution $f(J_1)$ of the difference equation (70). We define for this purpose the phase function

$$\Omega_0 = \sum_{i=1}^{3} (J_i \theta_i^0 + L_i \eta_i^0) \tag{79}$$

where

$$\theta_i^0 = \begin{cases} 0 & \text{if } 0 \le \theta_i \le \pi/2 \\ \pi & \text{if } \pi/2 < \theta_i \le \pi \end{cases}, \tag{80a}$$

$$\eta_i^0 = \begin{cases} 0 & \text{if } 0 \le \eta_i \le \pi/2 \\ \pi & \text{if } \pi/2 < \eta_i \le \pi \end{cases}, \tag{80b}$$

Ω_0 is either an integer or a half-integer multiple of π. It can be shown as for the phase function (49) defined for $3j$-coefficients that

$$\Omega_0/\pi \text{ half-integer (integer)} \Longleftrightarrow \Omega - \Omega_0 \text{ positive (negative)}. \tag{81}$$

We then set for the solution of (70)

$$f(J_1) = \cos\Omega_0 f_1 - \sin\Omega_0 f_2 \tag{82}$$

and it is readily checked that $f_1(J_1)$ and $f_2(J_1)$ thus defined must satisfy the modified difference equation

$$[\Delta^2(J_1) + 2 - 2\cos(\theta_1 - \theta_1^0)] f_i(J_1) \approx 0 \tag{83}$$

in the region of constant θ_1^0. For $J_{1\min} \ll J_1 \ll J_{1\max}$, $f(J_1)$ determined through (73) and (78) is

$$f(J_1) = (12\pi \sin\theta_1)^{-1/2} [\cos\Omega_0 \cos(\Omega - \Omega_0 + \alpha) - \sin\Omega_0 \sin(\Omega - \Omega_0 + \alpha)]. \tag{84}$$

Hence, $f_1(J_1)$ and $f_2(J_1)$ should be determined as to satisfy Eq. (83) in the neighborhood of $J_{1\min}$, $J_{1\max}$ and to match identically through Eq. (82) this expression for $J_{1\min} \ll J_1 \ll J_{1\max}$.

From the result of Sec. 3 one can infer that the solutions of the difference equation (83) in the semiclassical limit must obey the differential equation

$$\left(\frac{d^2}{dJ_1^2} - (\theta_1 - \theta_1^0)^2\right) \left(\frac{\sin(\theta_1 - \theta_1^0)}{\theta_1 - \theta_1^0}\right)^{1/2} f_i(J_1) = 0 \tag{85}$$

where J_1 is assumed to be a continuous variable. As was the case in Eq. (56), $(\theta_1 - \theta_1^0)^2$ is positive in the classical domain, zero at $J_{1\min}$, $J_{1\max}$ and negative in the nonclassical domain. The solutions of (85) and (56) are therefore formally identical, only the explicit algebraic form of the variables involved being different. Furthermore, the boundary conditions to be imposed on the solutions of the $6j$-coefficients (82) are identical to the boundary conditions postulated for the $3j$-coefficients, i.e., the $6j$-coefficients must decay to zero in the nonclassical domain as is illustrated by the example given in Fig. 3c. We have therefore for $J_{1\min} \le J_1 \le J_{1\max}$

$$f = \frac{Z^{1/4}}{(\sin\theta_1)^{1/2}} \times \begin{cases} a\cos\Omega_0 \operatorname{Ai}(-Z) - b\sin\Omega_0 \operatorname{Bi}(-Z), & \Omega - \Omega_0 < 0 \\ a'\cos\Omega_0 \operatorname{Bi}(-Z) - b'\sin\Omega_0 \operatorname{Ai}(-Z), & \Omega - \Omega_0 > 0 \end{cases} \tag{86}$$

and for $J_1 \le J_{1\min}$, $J_1 \ge J_{1\max}$

$$f = \frac{|Z|^{1/4}}{[\sinh|\theta_1 - \theta_1^0|]^{1/2}} \times \begin{cases} a\cos\Omega_0 \operatorname{Ai}(Z) - b\sin\Omega_0 \operatorname{Bi}(Z) & \text{for } \Omega - \Omega_0 < 0 \\ a'\cos\Omega_0 \operatorname{Bi}(Z) - b'\sin\Omega_0 \operatorname{Ai}(Z) & \text{for } \Omega - \Omega_0 > 0 \end{cases} \tag{87}$$

where

$$Z = (\tfrac{3}{2}|\Omega - \Omega_0|)^{2/3}. \tag{88}$$

Over the discrete set of quantum mechanically allowed J_i and L_i, the factor $\sin\Omega_0$ $(\Omega - \Omega_0 < 0)$ and $\cos\Omega_0$ $(\Omega - \Omega_0 > 0)$ of the irregular Airy functions $\operatorname{Bi}(Z)$ vanish in (87), so that (86) in the classical domain together with its analytical continuation (87) in the nonclassical domains indeed represent a possible solution. This solution matches identically with (84) if one chooses $a = b = a' = b' = S/\sqrt{12}$ in (86) and (87) and $\alpha = \pi/4$ in (84), S being an arbitrary phase factor. The $6j$-coefficients are then finally

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix} = S \frac{Z^{1/4}}{\sqrt{12V}} \times \begin{cases} \cos\Omega_0 \operatorname{Ai}(-Z) - \sin\Omega_0 \operatorname{Bi}(-Z), & \Omega - \Omega_0 < 0 \\ \cos\Omega_0 \operatorname{Bi}(-Z) - \sin\Omega_0 \operatorname{Ai}(-Z), & \Omega - \Omega_0 > 0 \end{cases} \tag{89}$$

in the classical domain and

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix} = S \frac{Z^{1/4}}{\sqrt{12|V|}} \times \begin{cases} \cos\Omega_0 \operatorname{Ai}(Z) - \sin\Omega_0 \operatorname{Bi}(Z), & \Omega - \Omega_0 < 0 \\ \cos\Omega_0 \operatorname{Bi}(Z) - \sin\Omega_0 \operatorname{Ai}(Z), & \Omega - \Omega_0 > 0 \end{cases} \tag{90}$$

in the nonclassical domain. The phase factor S may be determined according to the phase convention

$$\operatorname{sgn}\left(\begin{Bmatrix} j_{1\max} & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix}\right) = (-1)^{j_2+j_3+l_2+l_3}. \tag{91}$$

For $\Omega - \Omega_0 < 0$ one finds at $J_1 = J_{1\max}$

$$\begin{Bmatrix} \theta_1^0 & \theta_2^0 & \theta_3^0 \\ \eta_1^0 & \eta_2^0 & \eta_3^0 \end{Bmatrix} = \begin{Bmatrix} 0 & \pi & \pi \\ 0 & \pi & \pi \end{Bmatrix},$$

i.e., $\cos\Omega_0 = (-1)^{j_2+j_3+l_2+l_3}$. For $\Omega - \Omega_0 > 0$ there are two possibilities:

(a) $j_2 + j_3 > l_2 + l_3$:

$$\begin{Bmatrix} \theta_1^0 & \theta_2^0 & \theta_3^0 \\ \eta_1^0 & \eta_2^0 & \eta_3^0 \end{Bmatrix} = \begin{Bmatrix} \pi & \pi & \pi \\ 0 & 0 & 0 \end{Bmatrix}$$

(b) $j_2 + j_3 < l_2 + l_3$:

$$\begin{Bmatrix} \theta_1^0 & \theta_2^0 & \theta_3^0 \\ \eta_1^0 & \eta_2^0 & \eta_3^0 \end{Bmatrix} = \begin{Bmatrix} \pi & 0 & 0 \\ 0 & \pi & \pi \end{Bmatrix},$$

i.e., $-\sin\Omega_0 = (-1)^{j_1+j_2+j_3}$ or $-\sin\Omega_0 = (-1)^{j_1+l_2+l_3}$. But

$$\begin{Bmatrix} \theta_1^0 & \theta_2^0 & \theta_3^0 \\ \eta_1^0 & \eta_2^0 & \eta_3^0 \end{Bmatrix}$$

is constant throughout the nonclassical domains, hence for (a) $j_{1\max} = l_2 + l_3$, i.e., $-\sin\Omega_0 = (-1)^{j_2+j_3+l_2+l_3}$ and for (b) $j_{1\max} = j_2 + j_3$; i.e. $-\sin\Omega_0 = (-1)^{j_2+j_3+l_2+l_3}$. Since

in the nonclassical domain at $j_1 = j_{1\,\max}$ Ai(Z) and Bi(Z) are both positive, the sign of (90) is

$$(-1)^{j_2+j_3+l_2+l_3},$$

so that the sign convention (91) is met for $S=1$.

Let us finally compare our result (89) and (90) with the semiclassical expressions for the $6j$-coefficients stated by Ponzano and Regge. These authors give three formulas, one valid for $J_{1\,\min} \ll J_1 \ll J_{1\,\max}$, one for $J_1 \ll J_{1\,\min}$, $J_1 \gg J_{1\,\max}$ and one for $J_1 \approx J_{1\,\min}$, $J_{1\,\max}$. In the classical domain distant from $J_{1\,\min}$, $J_{1\,\max}$, we have

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix} \approx \frac{1}{\sqrt{12\pi V}} \cos\left(\Omega + \frac{\pi}{4}\right) \tag{92}$$

which is identical with the expression of Ponzano and Regge in this region. In the nonclassical domain distant from $J_{1\,\min}$, $J_{1\,\max}$ (90) becomes for J_i and L_i either integer or half-integer

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix} = (-1)^{j_2+j_3+l_2+l_3} \frac{1}{2\sqrt{12\pi\,|V|}} \exp(-|\Omega - \Omega_0|) \tag{93}$$

in agreement with the result of Ponzano and Regge.

To describe the $6j$-coefficients in the neighborhood of $J_{1\,\min}$, $J_{1\,\max}$ Ponzano and Regge introduced the variable

$$\frac{9}{2}\frac{V^3}{F_4} \approx |\Omega - \Omega_0| \tag{94}$$

where $F_4 = F(J_1, J_2, J_3)\, F(J_1, L_2, L_3)\, F(L_1, L_2, J_3)\, F(L_1, J_2, L_3)$ and the variable

$$\Phi = \begin{cases} \Omega_0 - 2\pi & \text{for } \Omega - \Omega_0 < 0 \\ \Omega_0 - \frac{3}{2}\pi & \text{for } \Omega - \Omega_0 > 0 \end{cases}.$$

Inserting these variables into (89) and (90) yields for $\Omega - \Omega_0 \gtrless 0$

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix} = 2^{-4/3} F_4^{-1/6} [\cos\Phi\ \mathrm{Ai}(-Z) \pm \sin\Phi\ \mathrm{Bi}(-Z)] \tag{89'}$$

and

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix} = 2^{-4/3} F_4^{-1/6} [\cos\Phi\ \mathrm{Ai}(Z) - \sin\Phi\ \mathrm{Bi}(Z)] \tag{90'}$$

where $Z = (3V)^2/(4F_4)^{2/3}$. For physical values of J_i and L_i, Φ is an integer multiple of π and (89′) and (90′) are then identical with the expressions of Ponzano and Regge.

Thus we have demonstrated that the semiclassical expressions for $6j$-coefficients first stated by Ponzano and Regge on the basis of only heuristic arguments can be derived systematically from the recursion relationship (66). It is admirable that Ponzano and Regge succeeded in obtaining their results without the guideline of a step-by-step derivation.

Equation (94) holds only in a small neighborhood of $J_{1\,\min}$, $J_{1\,\max}$. For this reason the expressions of Ponzano and Regge (89′) and (90′), though identical with our results (89) and (90) near $J_{1\,\min}$, $J_{1\,\max}$, do not match the functions (92) and (93) which hold distant from the classical boundaries. But our results (89) and (90) hold uniformly over the entire domain of quantum numbers and represent therefore an improvement over the approximations of Ponzano and Regge.

The recursion equation (66) of the $6j$-coefficients can be derived as the secular equation of a certain eigenvalue problem in which $l_1(l_1+1)$ represents the eigenvalues and $\nu_{l_1}(j_1) = \begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix}$ $(j_{1\,\min} \leq j_1 \leq j_{1\,\max})$ the eigenvectors.[3] We may have regarded $L_1^2 \approx l_1(l_1+1)$ in Eqs. (66c) and (67), (69) as unknown, carried through the derivation to be finally determined by the boundary conditions, i.e., $\nu_{l_1}(j_1) \to 0$ for $j_1 \gg J_{1\,\max}$ and $j_1 \ll J_{1\,\min}$. The semiclassical expressions (89) and (90) reveal then, that these boundary conditions are met only for L_1 being either integer or half-integer, for otherwise the coefficients $\sin\Omega_0(\Omega - \Omega_0 < 0)$ and $\cos\Omega_0(\Omega - \Omega_0 > 0)$ in (90) multiplying the irregular Airy functions do not vanish. The semiclassical solution of the $6j$-coefficient recursion (eigenvalue) equation exhibit thus the expected quantum character for L_1. More interestingly, it predicts the *exact* discrete set of quantum numbers (integer and half-integer). Since the solution (89) and (90) is independent of the recursion equation used as a starting point for the derivation (there are 6 different recursion equations), this remark holds as well for all variables in the $6j$-coefficients which must all be quantized according to Eq. (90). It is a remarkable fact that the semiclassical quantum numbers coincide with the set of exact quantum numbers for such a coincidence is only found in few situations (for example, for the Coulomb potential and harmonic oscillator eigenvalue problem) which possess special underlying symmetries. It may be speculated that it is the puzzling Regge-symmetry of $6j$-coefficients[10] (and $3j$-coefficients) which is responsible for this remarkable coincidence.

VI. COMPARISON OF EXACT AND SEMICLASSICAL WIGNER COEFFICIENTS

A comparison of exact and semiclassical $6j$-coefficients had already been carried out by Ponzano and Regge. Since earlier algorithms (and tables) for $6j$-coefficients were restricted to the domain of only moderate quantum numbers, these authors were not in a position to demonstrate directly the accuracy of semiclassical $6j$-coefficients (and $3j$-coefficients by the same token) involving large quantum numbers. Because of the surprisingly good agreement between exact and semiclassical $6j$-coefficients at small and moderate quantum numbers Ponzano and Regge expected that the semiclassical expressions should give very satisfactory values for large quantum number $6j$-coefficients.

Recently, we have developed an algorithm for the evaluation of $3j$- and $6j$-coefficients on the basis of the same recursion equations (5), (6), and (66) from which the semiclassical $3j$- and $6j$-coefficients had been derived as asymptotic solutions.[3] This algorithm was found numerically stable even for large quantum numbers-in fact, it served to evaluate the $3j$- and $6j$-coefficients in Figs. 1–3. Hence, it is now possible to examine directly the accuracy of large quantum number semiclassical Wigner coefficients. In Tables I–III are given some sample values of the Wigner coefficients

TABLE I. Accuracy of semiclassical $3j$-coefficients $\begin{pmatrix} j_1 & 100 & 60 \\ -10 & 60 & -50 \end{pmatrix}$.

j_1	Exact quantum mechanical[a]	Uniform semiclassical[b]
40	0.4999 (−04)	0.4833 (−04)
46	0.0848 (−01)	0.0847 (−01)
48	0.1563 (−01)	0.1562 (−01)
50	0.1857 (−01)	0.1858 (−01)
52	0.1139 (−01)	0.1141 (−01)
54	−0.0352 (−01)	−0.0349 (−01)
56	−0.1385 (−01)	−0.1385 (−01)
58	−0.0927 (−01)	−0.0929 (−01)
60	0.0519 (−01)	0.0517 (−01)
70	0.0119 (−01)	−0.0115 (−01)
80	−0.0213 (−01)	−0.0214 (−01)
90	−0.0808 (−01)	−0.0807 (−01)
100	−0.0926 (−01)	−0.0927 (−01)
106	0.0320 (−01)	0.0322 (−01)
108	0.1051 (−01)	0.1052 (−01)
110	0.1372 (−01)	0.1372 (−01)
112	0.1300 (−01)	0.1300 (−01)
114	0.1006 (−01)	0.1005 (−01)
116	0.0665 (−01)	0.0664 (−01)
118	0.0385 (−01)	0.0384 (−01)
120	0.0197 (−02)	0.0197 (−01)
130	0.1407 (−04)	0.1404 (−04)
140	0.7206 (−08)	0.7191 (−08)
150	0.1438 (−12)	0.1433 (−12)
160	0.3811 (−20)	0.3672 (−20)

[a] Evaluated by recursion, Ref. 3.
[b] Evaluated from Eqs. (60), (61).

TABLE II. Accuracy of semiclassical $3j$-coefficients $\begin{pmatrix} 120 & 60 & 70 \\ -10 & m & 10-m \end{pmatrix}$.

m	Exact quantum mechanical[a]	Uniform semiclassical[b]
−60	0.1749 (−30)	0.1626 (−30)
−50	0.7794 (−17)	0.7760 (−17)
−40	0.5450 (−09)	0.5436 (−09)
−30	0.4795 (−04)	0.4788 (−04)
−24	0.0286 (−01)	0.0286 (−01)
−22	0.0682 (−01)	0.0682 (−01)
−20	0.1207 (−01)	0.1207 (−01)
−18	0.1445 (−01)	0.1446 (−01)
−16	0.0841 (−01)	0.0842 (−01)
−14	−0.0462 (−01)	−0.0461 (−01)
−12	−0.1163 (−01)	−0.1163 (−01)
−10	−0.0255 (−01)	−0.0257 (−01)
0	0.0863 (−01)	0.0862 (−01)
10	0.0357 (−01)	0.0354 (−01)
20	−0.0810 (−01)	−0.0812 (−01)
22	−0.1043 (−01)	−0.1042 (−01)
24	0.0120 (−01)	0.0122 (−01)
26	0.1229 (−01)	0.1230 (−01)
28	0.1418 (−01)	0.1418 (−01)
30	0.0974 (−01)	0.0974 (−01)
32	0.0474 (−02)	0.0474 (−02)
34	0.0174 (−02)	0.0174 (−02)
40	0.1951 (−04)	0.1948 (−04)
50	0.8202 (−10)	0.8177 (−10)
60	0.3218 (−19)	0.3099 (−19)

[a] Evaluated by recursion, Ref. 3.
[b] Evaluated from Eqs. (60), (61).

presented in Figs. 1–3 together with the corresponding semiclassical values. The relative errors between the exact and the semiclassically evaluated $3j$- and $6j$-coefficients are found to be in general small (less than 1%). Exceptions are only the terminal $3j$- and $6j$-coefficients $\begin{pmatrix} j_{1\max} & 100 & 60 \\ -10 & 60 & -50 \end{pmatrix}$, etc. for which the error is of the order 1%.

We discussed above the connection between Eqs. (89), (90) and the semiclassical expressions (92), (93), (89′), (90′) stated by Ponzano and Regge. While (89), (90) provide a uniform approximation to the $6j$-coefficients over the entire domain of allowed quantum numbers, the expressions of Ponzano and Regge hold only over mu-

TABLE III. Accuracy of semiclassical $6j$-coefficients $\begin{Bmatrix} j_1 & 80 & 150 \\ 190 & 230 & 120 \end{Bmatrix}$.

j_1	Exact quantum mechanical[a]	Uniform semiclassical[b]	Ponzano and Regge semiclassical
110	0.3865 (−13)	0.3725 (−13)	0.3737 (−13)
120	0.2191 (−06)	0.2186 (−06)	0.2206 (−06)
126	0.0307 (−03)	0.0307 (−03)	0.0315 (−03)
128	0.0973 (−03)	0.0972 (−03)	0.1018 (−03)
130	0.2405 (−03)	0.2402 (−03)	0.2362 (−03)
132	0.4552 (−03)	0.4550 (−03)	0.4541 (−03)
134	0.6285 (−03)	0.6284 (−03)	0.6154 (−03)
136	0.5503 (−03)	0.5507 (−03)	0.5466 (−03)
138	0.1216 (−03)	0.1224 (−03)	0.1069 (−03)
140	−0.3852 (−03)	−0.3847 (−03)	−0.3931 (−03)
150	−0.3367 (−03)	−0.3359 (−03)	−0.3381 (−03)
160	−0.4230 (−03)	−0.4231 (−03)	−0.4231 (−03)
170	−0.3378 (−03)	−0.3372 (−03)	−0.3392 (−03)
180	−0.4400 (−03)	−0.4398 (−03)	−0.4491 (−03)
182	−0.0969 (−03)	−0.0963 (−03)	−0.1090 (−03)
184	0.3378 (−03)	0.3383 (−03)	0.3265 (−03)
186	0.5611 (−03)	0.5612 (−03)	0.5606 (−03)
188	0.5289 (−03)	0.5288 (−03)	0.5202 (−03)
190	0.3666 (−03)	0.3664 (−03)	0.3665 (−03)
192	0.2021 (−03)	0.2019 (−03)	0.1983 (−03)
194	0.0919 (−04)	0.0918 (−04)	0.0964 (−04)
200	0.3194 (−05)	0.3190 (−05)	0.3239 (−05)
210	0.5648 (−09)	0.5637 (−09)	0.5667 (−09)
220	0.1537 (−14)	0.1533 (−14)	0.1537 (−14)
230	0.2427 (−23)	0.2339 (−23)	0.2342 (−23)

[a] Evaluated by recursion, Ref. 3.
[b] Evaluated from Eqs. (89), (90).

TABLE IV. Convergence of semiclassical $3j$-coefficients $\begin{pmatrix} j_1\lambda & 4.5\lambda & 3.5\lambda \\ \lambda & -3.5\lambda & 2.5\lambda \end{pmatrix}$.

j_1	$\lambda=1$	$\lambda=2$	$\lambda=4$	$\lambda=8$	$\lambda=16$	$\lambda=32$	
1	0.2789 (00)	0.1769 (00)	0.9692 (−01)	0.4032 (−01)	0.0977 (−01)	0.0806 (−02)	QM[a]
	0.3043 (00)	0.1959 (00)	0.9520 (−01)	0.3878 (−01)	0.0926 (−01)	0.0758 (−02)	SC
2	−0.9535 (−01)	−0.8063 (−01)	0.0457 (−01)	−0.3174 (−01)	0.1288 (−01)	0.0297 (−02)	QM
	−0.9303 (−01)	−0.8294 (−01)	0.0523 (−01)	−0.3172 (−01)	0.1285 (−01)	0.0285 (−02)	SC
3	−0.6742 (−01)	−0.6995 (−01)	−0.0141 (−01)	−0.2728 (−01)	0.1378 (−01)	0.6773 (−02)	QM
	−0.6975 (−01)	−0.7031 (−01)	−0.0143 (−01)	−0.2730 (−01)	0.1378 (−01)	0.6773 (−02)	SC
4	0.1533 (00)	0.5646 (−01)	−0.2083 (−01)	−0.1316 (−01)	−0.0825 (−01)	−0.0275 (−02)	QM
	0.1558 (00)	0.5571 (−01)	−0.2158 (−01)	−0.1325 (−01)	−0.0823 (−01)	−0.0281 (−02)	SC
5	−0.1564 (00)	0.9203 (−01)	0.5313 (−01)	0.2863 (−01)	0.1021 (−01)	−0.4379 (−02)	QM
	−0.1566 (00)	0.9204 (−01)	0.5315 (−01)	0.2865 (−01)	0.1023 (−01)	−0.4376 (−02)	SC
6	0.1099 (00)	0.4867 (−01)	0.1711 (−01)	0.0397 (−01)	0.4149 (−03)	0.8876 (−05)	QM
	0.1090 (00)	0.4835 (−01)	0.1704 (−01)	0.0397 (−01)	0.4145 (−03)	0.8870 (−05)	SC
7	−0.5536 (−01)	0.1177 (−01)	0.9524 (−03)	0.1178 (−04)	0.3496 (−08)	0.6071 (−15)	QM
	−0.5441 (−01)	0.1162 (−01)	0.9452 (−03)	0.1173 (−04)	0.3488 (−08)	0.6064 (−15)	SC
8	0.1800 (−01)	0.9460 (−03)	0.4072 (−05)	0.1220 (−09)	0.1804 (−18)	0.6568 (−36)	QM
	0.1727 (−01)	0.9079 (−03)	0.3915 (−05)	0.1174 (−09)	0.1738 (−18)	0.6332 (−36)	SC

[a] QM = quantum mechanical values; SC = semiclassical values.

tually exclusive regions of the quantum number domain and do not connect smoothly with each other near the classical boundaries $J_{i\,\min}$, $J_{i\,\max}$ ($M_{i\,\min}$, $M_{i\,\max}$) In Tables I—III the $3j$- and $6j$-coefficient quantum mechanical and semiclassical values are presented for several consecutive values near the classical boundaries to demonstrate the uniformity of the semiclassical formulas (60), (61) and (89), (90). As can be seen from the quoted numerical values, the uniform expressions are very accurate and furnish thereby an improvement over the expressions of Ponzano and Regge.

The derivation of the semiclassical Wigner coefficients as the asymptotic solutions to the recursion equations (5), (6) and (66) had been based on the expansion of

$$\begin{pmatrix} \lambda j_1 & \lambda j_2 & \lambda j_3 \\ \lambda m_1 & \lambda m_2 & \lambda m_3 \end{pmatrix} \text{ and } \begin{Bmatrix} \lambda j_1 & \lambda j_2 & \lambda j_3 \\ \lambda l_1 & \lambda l_2 & \lambda l_3 \end{Bmatrix}$$

in terms of powers of λ, such that all terms up to order $O(\lambda^{-1})$ had been kept. This suggests that the semiclassical Wigner coefficients should converge to the quantum mechanical values with increasing λ. This conjecture is examined in Tables IV and V, for $3j$- and $6j$-coefficients. In Table IV the series of $3j$-coefficients $\begin{pmatrix} j_1\lambda & 4.5\lambda & 3.5\lambda \\ \lambda & -3.5\lambda & 2.5\lambda \end{pmatrix}$ are evaluated for $\lambda=1,2,4,8,16,32$ by means of recursion of Eq. (5) and by its asymptotic semiclassical solution (60), (61). As can be inferred from the tabulated values the relative error of the semiclassical $3j$-coefficients does decrease with increasing λ. Similarly, one can observe from Table V which presents the values of the $6j$-coefficients $\begin{Bmatrix} j_1\lambda & 8\lambda & 7\lambda \\ 6.5\lambda & 7.5\lambda & 7.5\lambda \end{Bmatrix}$ for $\lambda=1,2,4,8,16$ that the semiclassical $6j$-coefficients converge to the exact $6j$-coefficients with increasing λ.

Finally, the question may be raised if for very large quantum numbers a semiclassical evaluation of the

TABLE V. Convergence of semiclassical $6j$-coefficients $\begin{Bmatrix} j_1\lambda & 8\lambda & 7\lambda \\ 6.5\lambda & 7.5\lambda & 7.5\lambda \end{Bmatrix}$.

j_1	$\lambda=1$	$\lambda=2$	$\lambda=4$	$\lambda=8$	$\lambda=16$	
1	0.3491 (−01)	0.1218 (−01)	0.3226 (−02)	0.0513 (−02)	0.0302 (−03)	QM[a]
	0.3482 (−01)	0.1201 (−01)	0.3155 (−02)	0.0499 (−02)	0.0292 (−03)	SC
3	0.1891 (−01)	−0.7077 (−02)	0.0185 (−02)	−0.1458 (−02)	0.2156 (−03)	QM
	0.1905 (−01)	−0.7068 (−02)	0.0180 (−02)	−0.1458 (−02)	0.2157 (−03)	SC
5	−0.2359 (−01)	0.8663 (−02)	0.2973 (−02)	0.0887 (−02)	0.1358 (−03)	QM
	−0.2382 (−01)	0.8706 (−02)	0.2982 (−02)	0.0889 (−02)	0.1362 (−03)	SC
7	0.0129 (−01)	−0.7728 (−02)	0.1603 (−02)	−0.0698 (−02)	0.0315 (−03)	QM
	0.0152 (−01)	−0.7717 (−02)	0.1596 (−02)	−0.0699 (−02)	0.0318 (−03)	SC
9	0.1677 (−01)	0.0231 (−02)	−0.2800 (−02)	0.0854 (−02)	0.0697 (−03)	QM
	0.1671 (−01)	0.0198 (−02)	−0.2800 (−02)	0.0854 (−02)	0.0696 (−03)	SC
11	−0.2135 (−01)	0.7795 (−02)	0.2264 (−02)	−0.0020 (−02)	−0.3562 (−03)	QM
	−0.2147 (−01)	0.7793 (−02)	0.2259 (−02)	−0.0022 (−02)	−0.3561 (−03)	SC
13	0.2521 (−01)	0.9407 (−02)	0.1724 (−02)	−0.1184 (−02)	0.4040 (−03)	QM
	0.2527 (−01)	0.9429 (−02)	0.1731 (−02)	−0.1183 (−02)	0.4039 (−03)	SC
15	0.0271 (−01)	0.7636 (−04)	0.1171 (−06)	0.5415 (−12)	0.2293 (−22)	QM
	0.0257 (−01)	0.7175 (−04)	0.1095 (−06)	0.5051 (−12)	0.2136 (−22)	SC

[a] QM = quantum mechanical values; SC = semiclassical values.

Wigner coefficients should be favored over a recursive evaluation. The recursive evaluation of the Wigner coefficients generates simultaneously whole strings of $3j$- and $6j$-coefficients like

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \text{ and } \begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix} \text{ for all allowed } j_1,$$

whereas the semiclassical formulas need to evaluate each $3j$- and $6j$-coefficient individually. In cases for which such whole arrays of coefficients are needed, the recursive method involves less numerical effort than the semiclassical method, and also provides more accurate numerical values. In cases where only individual coupling coefficients are needed a semiclassical evaluation may nevertheless be quite useful. In addition to numerical evaluation, our systematic derivation of the semiclassical Wigner coefficients should contribute to a better physical understanding of the quantum mechanical theory of angular momentum coupling.

*Supported by the National Science Foundation.
†Current Address: Max-Planck-Institut für biophysikalische Chemie, Göttingen, Germany.

[1]G. Ponzano and T. Regge, in *Spectroscopy and Group Theoretical Methods in Physics* (North-Holland, Amsterdam, 1968).
[2]W. Miller, Adv. Chem. Phys. 25, 69 (1974).
[3]K. Schulten and R.G. Gordon, J. Math. Phys. 16, 1961 (1975), preceeding paper.
[4]E.U. Condon and G.H. Shortley, *The Theory of Atomic Spectra* (Cambridge U.P., Cambridge, 1935).
[5]E.P. Wigner, *Group Theory* (Academic, New York, 1959).
[6]G. Racah, Phys. Rev. 62, 438 (1942).
[7]A.R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, 1957).
[8]M. Abramowitz and I.A. Stegun, Appl. Math. Ser. 55 (1965).
[9]L. Schläfli, Quart. J. Math. 2, 269 (1858); 3, 54 (1860); 3, 97 (1860).
[10]T. Regge, Nuovo Cimento 10, 544 (1958); 11, 116 (1959).

An inverse problem in statistical mechanics*

S. Fesciyan and H. L. Frisch

Department of Chemistry, State University of New York at Albany, Albany, New York 12222
(Received 27 May 1975)

We study some one-dimensional systems specified by their nearest neighbor distribution. It is found that systems exhibiting second order phase transitions can be constructed. This paradoxical situation is resolved by a consideration of the peculiarities of the underlying potential. It is shown that the usual compressibility equation is not satisfied for these systems.

I. INTRODUCTION

The standard approach to classical statistical mechanics begins with the specification of an interaction potential among the particles. This determines a distribution defined on the phase space of a system at equilibrium in the manner of Gibbs. This distribution is then used in the calculation of ensemble averages of physical interest.

An "inverse problem" of statistical mechanics can be posed as follows: Given a distribution of particles what is the nature of the underlying potential which would make that distribution Gibbsian? Recently, we have examined this question in connection with a one-dimensional system where the questions of statistical mechanics are formulated and solved as problems in geometrical probability.[1] In Ref. 1 certain geometrical specifications defining the system were used to compute the nearest neighhor distribution whose asymptotic behavior determined the equation of state. The pressure $p(\rho)$ as a function of the density ρ was found to be a smooth curve defined up to a maximum density characteristic of the model. The underlying potential was found in Ref. 1 to be explicitly density dependent. In this paper we consider a more formal approach by assuming a nearest neighbor distribution $f(x, \lambda)$ depending on a parameter $\lambda \geq 0$, and show how it is possible to construct one-dimensional systems which behave in a physically reasonable manner and yet violate known results concerning one-dimensional classical systems. We interpret this in terms of the peculiarities of the underlying potential.

II. A SECOND ORDER PHASE TRANSITION

Consider a one-dimensional system with nearest neighbor distribution $f(x, \lambda)$ subject to the conditions

$$\int_0^\infty f(x, \lambda)\, dx = 1 \tag{1}$$

and

$$f(x, \lambda) \sim \exp[-p(\lambda)x] \quad \text{as } x \to \infty. \tag{2}$$

The density $\rho(\lambda)$ is given by

$$1/\rho(\lambda) = \int_0^\infty x f(x, \lambda)\, dx. \tag{3}$$

Eliminating λ between $\rho = \rho(\lambda)$ and $p = p(\lambda)$ gives us the equation of state

$$p = p(\rho). \tag{4}$$

If we further suppose that consecutive intervals are independently distributed, then the formula[2] [see also Eq. (11) in Ref. 1]

$$f(x, \lambda) = \frac{\exp[-p(\lambda)x]\exp[-V(x, \lambda)]}{\int_0^\infty \exp[-p(\lambda)x]\exp[-V(x, \lambda)]\, dx} \tag{5}$$

determines an underlying pair potential $V(x, \lambda)$ up to a function of λ. In general the interaction depends on λ as indicated. [$p(\lambda)$ denotes β times the pressure, and $V(x, \lambda)$ β times the pair potential throughout this paper with $\beta = 1/kT$.] Given $f(x, \lambda)$ correlations are also determined via the formula

$$\rho g(x) = \sum_{n=1}^{\infty} (f * f * \cdots * f)_n(x)$$

and

$$g(r_1 \cdots r_n) = \prod_{i<j} g(r_i - r_j). \tag{6}$$

The pair correlation function $g(x)$ is an infinite sum with each term denoting the n-fold convolution of $f(x, \lambda)$. The infinite sum is just the sum of nth-neighbor distributions, i. e., two particles can be either nearest neighbors or second neighbors, etc.

For the sake of illustration consider the following examples:

(a) The pure exponential case, $f(x, \lambda) = \lambda \exp(-\lambda x)$. This is of course the ideal gas with $p = \rho = \lambda$. Equation (6) gives $g(x) = 1$ readily.

(b) If $f(x, \lambda)$ is chosen to be an exponential outside a hard-core region

$$f(x, \lambda) = \begin{cases} 0, & x < \sigma, \\ \lambda \exp[-\lambda(x - \sigma)], & x > \sigma, \end{cases}$$

we obtain the case of a hard-rod gas. Equations (2) and (3) give

$$\rho(\lambda) = (1 + \lambda\sigma)/\lambda, \quad p(\lambda) = \lambda,$$

yielding $p(\rho) = \rho/1 - \rho\sigma$.

We now consider $f(x, \lambda)$ to be a linear combination of exponentials:

$$f(x, \lambda) = \tfrac{1}{2}\{p_1(\lambda)\exp[-p_1(\lambda)x] + p_2(\lambda)\exp[-p_2(\lambda)x]\} \tag{7}$$

with $p_1(\lambda)$ and $p_2(\lambda)$ monotone increasing functions of λ, and $p_1(\lambda) \to 0$ as $\lambda \to 0$. Equation (3) implies that

$$\frac{1}{\rho(\lambda)} = \frac{1}{2}\left(\frac{1}{p_1(\lambda)} + \frac{1}{p_2(\lambda)}\right). \tag{8}$$

We must have that $p_2(\lambda) \sim p_1(\lambda)$ as $\lambda \to 0$, so that the physical condition

$$p(\lambda) \sim \rho(\lambda) \quad \text{as } \rho \to 0 \tag{9}$$

is satisfied.

Clearly we can choose the functions $p_1(\lambda)$ and $p_2(\lambda)$ in a way to make this system exhibit a second order phase transition of the Ehrenfest type. Let $p_1(\lambda) \leq p_2(\lambda)$ up to $\lambda = \lambda_t$ and $p_2(\lambda) < p_1(\lambda)$ for $\lambda > \lambda_t$. Equations (8) and (9) and the monotonicity condition ensure that $d\rho/d\rho > 0$, and that the ideal gas law is obeyed as $\rho \to 0$.

The asymptotic behavior

$$f(x, \lambda) \sim \exp[-p(\lambda)x] \tag{10}$$

with

$$p(\lambda) = \begin{cases} p_1(\lambda), & \lambda < \lambda_t, \\ p_2(\lambda), & \lambda > \lambda_t, \end{cases}$$

implies that $p = p(\rho)$ has two branches with a discontinuous derivative at $\rho_t = \rho(\lambda_t)$. Choose, for example, $p_1(\lambda) = \lambda$, and $p_2(\lambda) = \lambda$ for $\lambda < \lambda_t$ and $p_2(\lambda) = (\lambda + \lambda_t)/2$ for $\lambda > \lambda_t$. The correlations in this model are given according to Eq. (6) by

$$\begin{aligned} \rho(\lambda) g(x, \lambda) = {} & \sum_{n=1}^{\infty} \frac{x^{n-1}}{(n-1)!} \left[p_1^n \exp(-p_1 x) + p_2^n \exp(-p_2 x) \right] \\ & + (-)^n \sum_{k=1}^{n-1} \binom{n}{k} \frac{p_1^{n-k} p_2^k}{(n-k-1)!\,(k-1)!} \frac{\partial^{n-k-1}}{\partial p_1^{n-k-1}} \frac{\partial^{k-1}}{\partial p_2^{k-1}} \\ & \times \left(\frac{\exp(-p_2 x) - \exp(-p_1 x)}{p_1 - p_2} \right). \end{aligned} \tag{11}$$

It remains to explain the rather paradoxical equation of state (10). The nonexistence of this type of behavior in one-dimensional systems has been shown under very general conditions.[3] For systems with nearest neighbor interactions the nonexistence of a phase transition is usually shown by the formula[4]

$$\begin{aligned} \frac{1}{\rho^2} \frac{\partial \rho}{\partial p} = {} & \frac{1}{2} \frac{1}{\varphi^2(\rho)} \int_0^{\infty} \int_0^{\infty} (x-y)^2 \exp[-p(x+y)] \\ & \times \exp\{-[V(x) + V(y)]\} dx\, dy \end{aligned} \tag{12}$$

where

$$\varphi(p) = \int_0^{\infty} \exp(-px) \exp[-V(x)]\, dx.$$

Boundedness and continuity in p of the right-hand side in (12) rules out both first and higher order phase transitions. However, if we allow for interactions $V(x; \rho)$ depending explicitly on the density the analogous formula would be

$$\begin{aligned} \frac{1}{\rho^2} \frac{\partial \rho}{\partial p} = {} & \frac{1}{2} \frac{1}{\varphi^2(p)} \int_0^{\infty} \int_0^{\infty} (x-y)^2 \exp[-p(x+y)] \\ & \times \exp\{-[V(x;\rho) + V(y;\rho)]\} \\ & - \frac{1}{\varphi^2(p)} \int_0^{\infty} dx \int_0^{\infty} dy \left[\left(\frac{\partial^2 V(x)}{\partial \rho^2} \right) \left(\frac{\partial \rho}{\partial p} \right)^2 \right. \\ & \left. + \left(\frac{\partial V(x)}{\partial \rho} \right) \left(\frac{\partial^2 \rho}{\partial p^2} \right) \right] \exp[-p(x+y)] \\ & \times \exp\{-[V(x;\rho) + V(y;\rho)]\}. \end{aligned} \tag{13}$$

The derivation of both (12) and (13) consists of differentiating the chemical potential $\mu = -\ln\varphi(p)$ with respect to p twice. We see that the derivatives of $\rho(p)$ are involved in both sides of the equation and, for example, a discontinuity in $\partial\rho/\partial p$ is no longer excluded. At a point of discontinuity of $\partial\rho/\partial p$ Eq. (13) would simply not be defined. If we allow $V(x; \rho)$ to be nonanalytic in ρ then we can construct systems with phase transitions of any order. This is precisely the situation that occurs with our choice of $f(x, \lambda)$ in (7). Formula (5) in conjunction with Eq. (10) clearly shows that the pair potential underlying this system must be density dependent and nonanalytic in ρ at $\rho = \rho_t$. This is because from (5) we have

$$\ln V = \ln\{\exp[p(\lambda)x] f(x, \lambda) N(\lambda)\}$$

where $N(\lambda)$ denotes the normalization in (5). Using (7)

$$\begin{aligned} \frac{\partial}{\partial x} \ln V \Big|_{x=0} &= p(\lambda) + \frac{\partial}{\partial x} \ln f(x, \lambda) \Big|_{x=0} \\ &= \begin{cases} \frac{1}{2} \frac{(p_1 p_2 - p_2^2)}{f(0, \lambda)}, & \rho < \rho_t, \\ \frac{1}{2} \frac{(p_1 p_2 - p_1^2)}{f(0, \lambda)}, & \rho > \rho_t, \end{cases} \end{aligned}$$

which is clearly density dependent. This fact removes the above-mentioned paradox.

III. CONCLUDING REMARKS

The one-dimensional systems specified by (7) do not satisfy the usual compressibility equation of equilibrium statistical mechanics

$$kT \frac{\partial \rho}{\partial p} = 1 + \rho \int [g(r) - 1]\, dr \tag{14}$$

because explicit density dependence of the potential contributes to the compressibility in a complicated way. A word about the thermodynamics of such systems is in order. From the general expression for the Gibbs free energy in one-dimension

$$G(p, \bar{N}) = -\bar{N} \ln \int_0^{\infty} \exp(-px) \exp[-V(x; p)]\, dx \tag{15}$$

with the nearest neighbor pair potential V dependent on macroscopic parameters as indicated, we can see that the thermodynamic variable conjugate to the pressure is not L, the size of the system, but a variable

$$\tilde{L} = L + \bar{N} f(p) \tag{16}$$

where

$$f(p) = \frac{\int_0^{\infty} [\partial V(x; p)/\partial p] \exp(-px) \exp[-V(x, p)]\, dx}{\int_0^{\infty} \exp(-px) \exp[-V(x, p)]\, dx}. \tag{17}$$

The Gibbs–Duhem relation for fixed temperature becomes

$$d\mu = \tilde{v}\, dp, \quad \tilde{v} = \tilde{L}/\bar{N} \equiv v + f(p). \tag{18}$$

It now follows that (as in p. 105 of Ref. 5)

$$-\frac{L}{(\bar{N})^2} \left(\frac{\partial \bar{N}}{\partial \mu} \right) = \frac{1}{v + f(p)} \left(\frac{\partial v}{\partial p} + \frac{\partial f}{\partial p} \right) \tag{19}$$

and following the derivation of Hill[5] we obtain

$$\begin{aligned} & -\frac{kT}{v + f(p)} \left(\frac{1}{v} \frac{\partial v}{\partial p} + \frac{1}{v} \frac{\partial f}{\partial p} \right) \\ & = 1 + \rho \int [g^{(2)}(r) - 1]\, dr - \frac{\rho}{2} \int dr \frac{\partial V(r; \mu)}{\partial \mu} g^{(2)}(r) \\ & \quad - \frac{\rho^2}{2} \iint dr_1\, dr_2 \frac{\partial}{\partial \mu} [V(r_1; \mu) + V(r_2; \mu) + V(r_1 + r_2; \mu)] \\ & \quad \times g^{(3)}(r_1, r_2) \end{aligned} \tag{20}$$

which reduces to the usual compressibility relation for $f(p) = 0$. The explicit form of the variable conjugate to p in three dimensions is not apparent; however, the generalized compressibility equation in three dimensions should be essentially the same as (20) with a suitable $f(p)$ which must be determined. Equation (14) is satisfied by the pair correlation function of the usual systems considered in statistical physics, which, even if forces are not derivable from a density independent pairwise additive potential, are supposed to possess forces derivable from a more general density-independent many body potential.[6] The system considered in Sec. II cannot correspond to a system with density independent short range many-body potential. The latter will not exhibit a one-dimensional phase transition.

An extension of the inverse approach presented to more than one dimension is not straightforward. To obtain a consistent set of correlation functions the best resort seems to be a geometrical method like the one considered in Ref. 1. A system thus defined will in general correspond to one with density dependent many-body potentials defined again in purely statistical terms, i. e., in terms of

$$P(x_1, \ldots, x_N) \propto \exp[-U(x_1, \ldots, x_N)],$$

the probability that a given region of volume V will contain N particles in the configuration $(x_1, \ldots, x_N)$. Such a system will also not obey the compressibility Eq. (14).

*Research supported by NSF Grant No. GP42614X.

[1]S. Fesciyan and H. L. Frisch, J. Stat. Phys. **12**, 153 (1975).
[2]M. Wertheim, in *Classical Fluids*, edited by H. L. Frisch and J. L. Lebowitz (Benjamin, New York, 1964), p. II. 279.
[3]R. L. Dobrushin, Commun. Math. Phys. **32**, 269 (1973); D. Ruelle, Commun. Math. Phys. **2**, 267 (1968); G. Gallavotti, S. Miracle-Sole, and D. Ruelle, Phys. Lett. **26A**, 350 (1968).
[4]I. Z. Fisher, *Statistical Theory of Liquids* (University of Chicago Press, Chicago, 1964), p. 36.
[5]T. L. Hill, *Statistical Mechanics* (McGraw-Hill, New York, 1956)
[6]F. H. Stillinger, J. Chem. Phys. **57**, 1780 (1972).

Direct determination of the Iwasawa decomposition for noncompact semisimple Lie algebras

J. F. Cornwell

Department of Theoretical Physics, University of St. Andrews, St. Andrews, Fife, Scotland
(Received 3 December 1974)

A direct method for the determination of the Iwasawa decomposition of any noncompact semisimple real Lie algebra is described in detail. It is based on the canonical form of the Lie algebra. The physically important Lie algebras $so(3,1)$, $so(4,1)$, $so(3,2)$, and $so(4,2)$ are treated as illustrative examples.

I. INTRODUCTION

The Iwasawa[1] decomposition plays a key role in the construction of the infinite-dimensional unitary irreducible representations of the noncompact semisimple Lie groups by the "induced representation" method, accounts of which may be found for example in the reviews of Mackey,[2,3] Bruhat,[4] and Stein.[5]

The usual method for the determination of the Iwasawa decomposition of a noncompact semisimple Lie algebra involves solving certain simultaneous eigenvalue type equations. A description of this approach may be found in the book of Hermann,[6] who also gives some examples, and a further typical application [to the Lie algebra $so(4,2) \sim su(2,2)$] appears in the work of Kihlberg, Müller, and Halbwachs.[7] However, this method is essentially indirect, for it involves a totally unnecessary redetermination of the structure of the Lie algebra.

The present paper describes a much more simple and direct method for obtaining the Iwasawa decomposition. It is based on the canonical form of the structure of a noncompact semisimple real Lie algebra as it is generated through Cartan's theorem[8,9] by the application of an involutive automorphism to the compact real form. As is well known, it is particularly convenient to use in this process the *chief* involutive automorphisms introduced by Gantmacher.[10,11] Although its main virtue is its directness, the method also has the advantage of producing the decomposition immediately in the canonical form, which is the most convenient for many applications, particularly those involving semisimple subalgebras.[12-17] [If, however, another form of the structure is needed in some context, as, for example, in discussions of nonsemisimple subalgebras,[18,19] it is easy to construct the mapping from the canonical form. In particular, for the Lie algebras $so(p,q)$ there is a simple mapping[20] from the canonical form to the form consisting of $(p+q)\times(p+q)$ matrices $\mathbf{a}$ such that $\mathrm{tr}\mathbf{a}=0$ and $\tilde{\mathbf{a}}\mathbf{g}+\mathbf{g}\mathbf{a}=0$, where $\mathbf{g}$ is a diagonal matrix with p elements $+1$ and q elements -1.]

In Sec. II an account is given of the standard theory of the Iwasawa decomposition, the direct method of determination being described in Sec. III. This involves an automorphism V between the two Cartan subalgebras that appear. This automorphism V is examined in Sec. IV for the case in which the noncompact real Lie algebra is generated by an inner involutive automorphism, the generalization to outer involutive automorphisms being considered in Sec. V. The physically important Lie algebras $so(3,1)$, $so(4,1)$, $so(3,2)$, and $so(4,2)$ are treated as illustrative examples.

II. THE IWASAWA DECOMPOSITION

The essential feature of the Iwasawa decomposition in a canonical basis is that it involves *two* Cartan subalgebras $\mathcal{H}$ and $\mathcal{H}'$, which are not identical in general.

The first of these arises in the construction of the noncompact semisimple real Lie algebra $\mathcal{L}$ itself. It may be supposed that $\mathcal{L}$ is generated from its compact real form $\mathcal{L}_c$ by a *chief* involutive automorphism Z,[10-12] this chief automorphism being defined with respect to the Cartan subalgebra $\mathcal{H}$ of $\tilde{\mathcal{L}}$, $\tilde{\mathcal{L}}$ being the complexification of $\mathcal{L}$. That is, $\mathcal{L}=\sqrt{Z}\mathcal{L}_c$, where $\sqrt{Z}=\frac{1}{2}(1-i)Z+\frac{1}{2}(1+i)I$ (I being the identity), and where for a chief inner automorphism $Z=\exp(\mathrm{ad}h)$, where $h\in\mathcal{H}$,[10-12] while for a chief outer automorphism $Z=Z_0\exp(\mathrm{ad}h)$, $h\in\mathcal{H}$. Here Z_0 is an outer automorphism associated with a "rotation" in $\mathcal{H}$.[10,11,13]

The conventions and notations of Jacobson[21] will be used for the canonical form of $\tilde{\mathcal{L}}$. In particular, a typical root defined with respect to $\mathcal{H}$ will be denoted by $\alpha(h)$, $h\in\mathcal{H}$, and e_α, its corresponding element in $\mathcal{L}$, will be assumed to be such that

$$[e_\alpha, h]=\alpha(h)e_\alpha \tag{1}$$

for all $h\in\mathcal{H}$. For each such root α there exists an element $h_\alpha\in\mathcal{H}$ such that

$$B(h,h_\alpha)=\alpha(h) \tag{2}$$

for all $h\in\mathcal{H}$, $B(x,y)$ being the Killing form of $\tilde{\mathcal{L}}$. It may be assumed that the e_α are normalized so that

$$B(e_\alpha, e_{-\alpha})=-1, \tag{3}$$

which implies that

$$[e_\alpha, e_{-\alpha}]=h_\alpha. \tag{4}$$

A further useful bilinear form is defined by $(\alpha,\beta)=B(h_\alpha,h_\beta)$. The remaining commutation relations of $\tilde{\mathcal{L}}$ are

$$[e_\alpha, e_\beta]=N_{\alpha,\beta}e_{\alpha+\beta}, \tag{5}$$

where $N_{\alpha,\beta}=0$ if $\alpha+\beta$ is not a root of $\tilde{\mathcal{L}}$. Clearly, $N_{\beta,\alpha}=-N_{\alpha,\beta}$, and by convention[21] $N_{-\alpha,-\beta}=N_{\alpha,\beta}$. Moreover, if the α string of roots containing β is $\beta-r\alpha$, $\beta-(r-1)\alpha,\ldots,\beta,\ldots,\beta+q\alpha$, then the magnitude of $N_{\alpha,\beta}$ is given by $(N_{\alpha,\beta})^2=\frac{1}{2}q(r+1)(\alpha,\alpha)$, but the signs of the $N_{\alpha,\beta}$ are to some extent arbitrary. It will be assumed

that the rank of $\tilde{\mathcal{L}}$ is l, and that $\alpha_1, \alpha_2, \ldots, \alpha_l$ are the simple roots of $\tilde{\mathcal{L}}$ defined with respect to $\mathcal{H}$. The canonical form of the basis of the compact real form $\mathcal{L}_c$ may then be taken to consist of ih_α, for $\alpha = \alpha_1, \alpha_2, \ldots, \alpha_l$, together with $(e_\alpha + e_{-\alpha})$ and $i(e_\alpha - e_{-\alpha})$ for every root α of $\tilde{\mathcal{L}}$ defined with respect to $\mathcal{H}$.

The following account of the Iwasawa decomposition[1] is based on the exposition of Harish-Chandra,[22] although the notation is quite different. Let $\mathcal{P}$ be the maximal compact subalgebra of $\mathcal{L}$ defined such that $a \in \mathcal{K}$ if and only if $a \in \mathcal{L}$ and $Za = a$. Let $\mathcal{P}$ be the subspace of $\mathcal{L}$ such that $a \in \mathcal{P}$ if and only if $a \in \mathcal{L}$ and $Za = -a$. ($\mathcal{P}$ is not a subalgebra of $\mathcal{L}$.) Suppose that $\mathcal{A}$ is a maximal Abelian subalgebra of $\mathcal{P}$, that $\dim\mathcal{A} = m$, where $m \leq l$, and that $H'_1, H'_2, \ldots, H'_m$ form a basis for $\mathcal{A}$. Let $\mathcal{M}$ be a maximal Abelian subalgebra of $\mathcal{K}$ that commutes with $\mathcal{A}$. Then Iwasawa[1] showed that the complexification of $\mathcal{A} \oplus \mathcal{M}$ is a Cartan subalgebra of $\tilde{\mathcal{L}}$. This second Cartan subalgebra will be denoted by $\mathcal{H}'$. As $\dim\mathcal{H}' = l$, it follows that $\dim\mathcal{M} = l - m$. If $\dim\mathcal{M} > 0$, $\mathcal{M}$ may be taken to have basis elements $-iH'_{m+1}, -iH'_{m+2}, \ldots, -iH'_l$, so that $H'_1, H'_2, \ldots, H'_l$ form a basis for $\mathcal{H}'$. (It should be noted that apart from the above requirements the choice of $H'_1, \ldots, H'_l$ is completely arbitrary. In particular, they need not be associated in any way with the roots of $\tilde{\mathcal{L}}$ defined with respect to $\mathcal{H}'$.)

The final stage in the construction of the Iwasawa decomposition involves the determination of a nilpotent subalgebra $\mathcal{N}$ of $\mathcal{L}$ whose elements are linear combinations of those of both $\mathcal{K}$ and $\mathcal{P}$. This requires the roots $\alpha'(h')$ of $\tilde{\mathcal{L}}$ defined with respect to $\mathcal{H}'$ and the corresponding elements $e'_{\alpha'}$ of $\tilde{\mathcal{L}}$ defined such that

$$[e'_{\alpha'}, h'] = \alpha'(h') \tag{6}$$

(for all $h' \in \mathcal{H}'$) and

$$B(e'_{\alpha'}, e'_{-\alpha'}) = -1. \tag{7}$$

Moreover, let $h'_{\alpha'} \in \mathcal{H}'$ be such that

$$B(h', h'_{\alpha'}) = \alpha'(h') \tag{8}$$

for all $h' \in \mathcal{H}'$. Then

$$h'_{\alpha'} = \sum_{j=1}^{l} c_j(\alpha') H'_j \tag{9}$$

[where the coefficients $c_j(\alpha')$ are all real[22]], and the root α' (defined with respect to $\mathcal{H}'$) is said to be a member of the set P' if and only if $c_j(\alpha') > 0$, where j is the least index such that $c_j(\alpha') \neq 0$. (That is P' is the set of "positive" roots of $\tilde{\mathcal{L}}$ with respect to $\mathcal{H}'$.) The set P' may be divided into two disjoint subsets P'_- and P'_+ by the requirement that $\alpha' \in P'_-$ if and only if $\alpha' \in P'$ and

$$\alpha'(Zh') = \alpha'(h') \tag{10}$$

for all $h' \in \mathcal{H}'$, and by the requirement that $\alpha' \in P'_+$ if and only if $\alpha' \in P'$ and $\alpha' \notin P'_-$. If $\tilde{\mathcal{N}}$ is the subspace of $\tilde{\mathcal{L}}$ spanned by the elements $e'_{\alpha'}$ for all $\alpha' \in P'_+$, then Iwasawa[1] has shown that $\tilde{\mathcal{N}}$ is a nilpotent subalgebra of $\tilde{\mathcal{L}}$. The required subalgebra $\mathcal{N}$ of $\mathcal{L}$ is finally defined by $\mathcal{N} = \tilde{\mathcal{N}} \cap \mathcal{L}$. The Iwasawa decomposition of $\mathcal{L}$ is then

$$\mathcal{L} = \mathcal{K} \oplus \mathcal{A} \oplus \mathcal{N},$$

where $\oplus$ denotes here that $\mathcal{L}$ is the vector-space direct sum of $\mathcal{K}, \mathcal{A}$, and $\mathcal{N}$, and does not imply that $\mathcal{K}$, $\mathcal{A}$, and $\mathcal{N}$ mutually commute. The subalgebra of $\mathcal{L}$ that plays such a vital role in the induced representation method[2-5] is $\mathcal{M} \oplus \mathcal{A} \oplus \mathcal{N}$.

It will be clear that whereas the determination of $\mathcal{K}$, $\mathcal{A}$, and $\mathcal{M}$ is completely straightforward, the construction on $\mathcal{N}$ presents more difficulties. The usual method[6,7] consists of treating Eq. (6), or more precisely

$$[e'_{\alpha'}, H'_j] = \alpha'(H'_j) e'_{\alpha'}, \quad j = 1, \ldots, l, \tag{11}$$

as a set of l simultaneous eigenvalue equations in order to determine the "eigenvalues" $\alpha'(H'_j)$ and the "eigenvectors" $e'_{\alpha'}$. As will be shown in Sec. III, this is totally unnecessary, for the $e'_{\alpha'}$ are simply automorphic images of the e_α.

III. DIRECT DETERMINATION OF THE IWASAWA DECOMPOSITION

The direct determination of the Iwasawa decomposition is based on the fact that there exists an inner automorphism V that maps $\mathcal{H}'$ into $\mathcal{H}$. (In fact any two Cartan subalgebras of $\tilde{\mathcal{L}}$ are related by an inner automorphism.[10]) Suppose that $Vh' = h$, where $h' \in \mathcal{H}'$ and $h \in \mathcal{H}$. On applying V to Eq. (6) one obtains

$$[Ve'_{\alpha'}, h] = \alpha'(V^{-1}h)(Ve'_{\alpha'}),$$

which implies that $\alpha'(V^{-1}h)$ is a root of $\tilde{\mathcal{L}}$ with respect to $\mathcal{H}$ and $Ve'_{\alpha'}$ is the corresponding element of $\tilde{\mathcal{L}}$. Let $\alpha'(V^{-1}h)$ be denoted by $\alpha(h)$; that is,

$$\alpha(h) = \alpha'(V^{-1}h). \tag{12}$$

Equation (12) establishes a one-to-one correspondence between the roots defined with respect to $\mathcal{H}$ and the roots defined with respect to $\mathcal{H}'$. Moreover, as the Killing form is invariant under automorphisms, Eqs. (3) and (7) imply that

$$Ve'_{\alpha'} = e_\alpha,$$

and conversely

$$e'_{\alpha'} = V^{-1} e_\alpha.$$

Similarly Eqs. (2) and (8) imply that

$$Vh'_{\alpha'} = h_\alpha.$$

Now suppose that $H_1, H_2, \ldots, H_l$ of $\mathcal{H}$ are defined by

$$H_j = VH'_j, \quad j = 1, \ldots, l.$$

Then Eq. (9) becomes

$$h_\alpha = \sum_{j=1}^{l} c_j(\alpha') H_j. \tag{13}$$

It is now possible to define a set P of roots α defined with respect to $\mathcal{H}$ by the requirement that $\alpha \in P$ if and only if the corresponding root α' defined with respect to $\mathcal{H}'$ is a member of P'. With the further definition that $b_j(\alpha) = c_j(\alpha')$, $j = 1, \ldots, l$, Eq. (13) becomes

$$h_\alpha = \sum_{j=1}^{l} b_j(\alpha) H_j. \tag{14}$$

Thus

$\alpha \in P$ if and only if $b_j(\alpha) > 0$, where j is the least index such that $b_j(\alpha) \neq 0$. (A)

In the same way one may define the sets P_- and P_+ of

roots α defined with respect to $\mathcal{H}$ by the requirement that $\alpha \in P_-$ or P_+ if and only if the corresponding root α' defined with respect to $\mathcal{H}'$ is a member of P'_- or P'_+, respectively. Eqs. (10) and (12) imply that $\alpha \in P_-$ if and only if $\alpha \in P$ and $\alpha(VZV^{-1}h) = \alpha(h)$ for all $h \in \mathcal{H}$. However, this criterion can be further refined, for $H_1, H_2, \ldots, H_l$ form a basis for $\mathcal{H}$ and $VZV^{-1}H_j = H_j$ for $j = m+1, m+2, \ldots, l$, but $VZV^{-1}H_j = -H_j$ for $j = 1, 2, \ldots, m$. Thus

$$\alpha \in P_- \text{ if and only if } \alpha \in P \text{ and } \alpha(H_j) = 0 \text{ for } j = 1, 2, \ldots, m. \tag{B}$$

Also,

$$\alpha \in P_+ \text{ if and only if } \alpha \in P \text{ and } \alpha \notin P_-. \tag{C}$$

The subalgebra $\tilde{\mathcal{N}}$ is then easily characterized as being the subalgebra spanned by the elements $V^{-1}e_\alpha$ for all $\alpha \in P_+$.

To summarize, the procedure for determining $\mathcal{A}$, $\mathcal{M}$, and $\mathcal{N}$ consists of the following six stages:

(1) Choose a basis $H'_1, \ldots, H'_m$ of $\mathcal{A}$ and a basis $-iH'_{m+1}, \ldots, -iH'_l$ of $\mathcal{M}$.

(2) Find the inner automorphism V. (This step is discussed further in Secs. IV and V.)

(3) Evaluate $H_j = VH'_j$, $j = 1, \ldots, l$.

(4) Determine the sets P, P_-, and P_+ by criteria (A), (B), and (C), respectively.

(5) Evaluate $V^{-1}e_\alpha$ for all $\alpha \in P_+$, these being the basis elements of $\tilde{\mathcal{N}}$.

(6) Determine $\mathcal{N}$ by the definition $\mathcal{N} = \tilde{\mathcal{N}} \cap \mathcal{L}$.

IV. IWASAWA DECOMPOSITION FOR REAL LIE ALGEBRAS GENERATED BY CHIEF INNER INVOLUTIVE AUTOMORPHISMS

A. General theory

The chief inner involutive automorphism $Z = \exp(\mathrm{ad}h)$ is diagonal with respect to the canonical basis of $\mathcal{L}_c$ described in Sec. II. The basis elements ih_α, for $\alpha = \alpha_j$, $j = 1, \ldots, l$, all correspond to eigenvalue $+1$, while $(e_\alpha + e_{-\alpha})$ and $i(e_\alpha - e_{-\alpha})$ both correspond to eigenvalue $\exp\{\alpha(h)\}$ $(= \pm 1)$. [Expressions for h and for $\exp\{\alpha_j(h)\}$, $j = 1, \ldots, l$, for every classical simple real Lie algebra generated by an inner involutive automorphism are given in Ref. 12, the corresponding expressions for the exceptional Lie algebras being given in Ref. 16.]

$\mathcal{K}$ therefore has a basis consisting of ih_α, for $\alpha = \alpha_j$, $j = 1, \ldots, l$, together with $(e_\alpha + e_{-\alpha})$ and $i(e_\alpha - e_{-\alpha})$ for all α such that $\exp\{\alpha(h)\} = 1$, while the basis of $\mathcal{P}$ consists of $i(e_\alpha + e_{-\alpha})$ and $(e_\alpha - e_{-\alpha})$ for all α such that $\exp\{\alpha(h)\} = -1$. Thus $\mathcal{A}$ may be taken to have a basis consisting of elements of the form $i(e_\alpha + e_{-\alpha})$. Let $R_{\mathcal{A}}$ denote the set of positive roots α that appear this way in $\mathcal{A}$. In order to ensure that $\mathcal{A}$ is Abelian, $R_{\mathcal{A}}$ must be chosen so that if β and γ are any two roots in $R_{\mathcal{A}}$ then neither $\beta + \gamma$ nor $\beta - \gamma$ is a root of $\mathcal{L}$. Similarly, $\mathcal{M}$ may be taken to have a basis consisting of elements of the form $(e_\alpha + e_{-\alpha})$ (with the set of positive roots α appearing this way in $\mathcal{M}$ being denoted by $R_{\mathcal{M}}$), together possibly with some elements of $\mathcal{H} \cap \mathcal{L}$. $R_{\mathcal{M}}$ is characterized by the requirement that for any two roots β and γ of $R_{\mathcal{A}} \cup R_{\mathcal{M}}$ neither $\beta + \gamma$ nor $\beta - \gamma$ is a root of $\mathcal{L}$, while if $h'' \in \mathcal{H} \cap \mathcal{L}$ is a member of $\mathcal{M}$ then $\alpha(h'') = 0$ for all $\alpha \in R_{\mathcal{A}} \cup R_{\mathcal{M}}$.

The required inner automorphism V that maps $\mathcal{H}'$ into $\mathcal{H}$ is then given by

$$V = \prod_\alpha V_\alpha, \tag{15}$$

where

$$V_\alpha = \exp[\mathrm{ad}\{ia_\alpha(e_\alpha - e_{-\alpha})\}],$$

$$a_\alpha = \pi/\{8(\alpha, \alpha)\}^{1/2},$$

and the product in (15) is over all roots α of $R_{\mathcal{A}} \cup R_{\mathcal{M}}$. This may be seen as follows. Firstly, if $H'_j = i(e_\alpha + e_{-\alpha})$, $1 \le j \le m$, is a basis element of $\mathcal{A}$ chosen as above, then

$$V_\alpha H'_j = -\{2/(\alpha, \alpha)\}^{1/2} h_\alpha$$

[Eq. (A2) of the Appendix], which is certainly in $\mathcal{H}$. Moreover, if $\beta \in R_{\mathcal{A}} \cup R_{\mathcal{M}}$ and $\beta \neq \alpha$, then V_α commutes with V_β and $V_\beta h_\alpha = h_\alpha$. Thus

$$H_j = VH'_j = -\{2/(\alpha, \alpha)\}^{1/2} h_\alpha, \tag{16}$$

so that V maps $\mathcal{A}$ into $\mathcal{H}$. The same is immediately true for the subspace of $\mathcal{M}$ spanned by the basis elements $-iH'_j = e_\alpha + e_{-\alpha}$, as again $H'_j = i(e_\alpha + e_{-\alpha})$. It remains only to consider the subspace of $\mathcal{M}$ spanned by basis elements in $\mathcal{H} \cap \mathcal{L}$. By the above construction every V_α in V leaves these elements invariant, so that if $-iH'_j = ih''$, then

$$H_j = H'_j = -h''. \tag{17}$$

All the properties of the inner automorphisms V_α and their inverses V_α^{-1} that are needed for stage (5) of the

TABLE I. Dimensions of $\mathcal{L}$, $\mathcal{K}$, $\mathcal{A}$, $\mathcal{M}$, and $\mathcal{N}$ for the real forms $\mathcal{L}$ of $\tilde{\mathcal{L}} = A_l$, $l \ge 1$.

$\mathcal{L}$	$su(l+1-p, p)$, $0 \le p \le \frac{1}{2}(l+1)$	$sl(l+1, R)$	$q(l+1)/2$ l odd
$\dim\mathcal{L}$	$l^2 + 2l$	$l^2 + 2l$	$l^2 + 2l$
$\dim\mathcal{K}$	$l^2 + 2l - 2lp - 2p + 2p^2$	$\frac{1}{2}l(l+1)$	$\frac{1}{2}(l+1)(l+2)$
$\dim\mathcal{A}$	p	l	$\frac{1}{2}(l-1)$
$\dim\mathcal{M}$	$l-p$	0	$\frac{1}{2}(l+1)$
$\dim\mathcal{N}$	$p(2l+1-2p)$	$\frac{1}{2}l(l+1)$	$\frac{1}{2}(l^2-1)$

TABLE II. Dimensions of $\mathcal{L}$, $\mathcal{K}$, $\mathcal{A}$, $\mathcal{M}$, and $\mathcal{N}$ for the real forms $\mathcal{L}$ of $\tilde{\mathcal{L}} = B_l$, $l \ge 1$.

$\mathcal{L}$	$so(2l+1-2p, 2p)$, $o \le p \le \frac{1}{2}l$	$so(2l+1-2p, 2p)$, $\frac{1}{2}l < p \le l$
$\dim\mathcal{L}$	$2l^2 + l$	$2l^2 + l$
$\dim\mathcal{K}$	$2l^2 + l - 4lp + 4p^2 - 2p$	$2l^2 + l - 4lp + 4p^2 - 2p$
$\dim\mathcal{A}$	$2p$	$2(l-p) + 1$
$\dim\mathcal{M}$	$l - 2p$	$2p - l - 1$
$\dim\mathcal{N}$	$4p(l-p)$	$4lp + 4p - 4p^2 - 2l - 1$

TABLE III. Dimensions of $\mathcal{L}$, $\mathcal{K}$, $\mathcal{A}$, $\mathcal{M}$, and $\mathcal{N}$ for the real forms $\mathcal{L}$ of $\tilde{\mathcal{L}} = C_l$, $l \ge 1$.

$\mathcal{L}$	nsp_{21}^{2p}, $o \le p \le \frac{1}{2}l$	$sp(l, R)$
$\dim\mathcal{L}$	$2l^2 + l$	$2l^2 + l$
$\dim\mathcal{K}$	$2l^2 + l - 4lp + 4p^2$	l^2
$\dim\mathcal{A}$	p	l
$\dim\mathcal{M}$	$l - p$	o
$\dim\mathcal{N}$	$p(4l - 4p - 1)$	l^2

TABLE IV. Dimensions of $\mathcal{L}$, $\mathcal{K}$, $\mathcal{A}$, $\mathcal{M}$, and $\mathcal{N}$ for the real forms $\mathcal{L}$ of $\tilde{\mathcal{L}} = D_l$, $l \geq 2$.

$\mathcal{L}$	$so(2l-2p, 2p)$, $o \leq p \leq \frac{1}{2}l$	$so(2l-2p-1, 2p+1)$, $o \leq p < \frac{1}{2}l$	ND_{2l}, l odd	ND_{2l}, l even
$\dim\mathcal{L}$	$2l^2+l$	$2l^2+l$	$2l^2+l$	$2l^2+l$
$\dim\mathcal{K}$	$2l^2-l-4lp+4p^2$	$2l^2-3l+4p^2-4lp+1+4p$	l^2	l^2
$\dim\mathcal{A}$	$2p$	$2p+1$	$\frac{1}{2}(l-1)$	$\frac{1}{2}l$
$\dim\mathcal{M}$	$l-2p$	$l-2p-1$	$\frac{1}{2}(l+1)$	$\frac{1}{2}l$
$\dim\mathcal{N}$	$2p(2l-2p-1)$	$2l-4p^2+4lp-2-6p$	$\frac{1}{2}(l-1)(2l-1)$	$\frac{1}{2}l(2l-3)$

procedure of Sec. III are summarized in the Appendix. It should be noted that if $\alpha \in R_{\mathcal{A}} \cup R_{\mathcal{M}}$ then Eq. (16) implies that $-\alpha \in P_+$, so that [by Eq. (A4)] $V^{-1}e_{-\alpha} = -\frac{1}{2}(e_\alpha - e_{-\alpha}) - i\{2/(\alpha,\alpha)\}^{1/2}h_\alpha$ is a basis element of both $\bar{\mathcal{N}}$ and $\mathcal{N}$.

The dimensions of $\mathcal{K}$, $\mathcal{A}$, $\mathcal{M}$, and $\mathcal{N}$ for all the simple classical Lie algebras are listed in Tables I to V. These dimensions are easily calculated. In particular, $\dim\mathcal{K} = \frac{1}{2}\{\dim\mathcal{L} - \delta\}$, where δ is the character of $\mathcal{L}$,[8,9,11,12] $\dim\mathcal{M} = l - \dim\mathcal{A}$, and $\dim\mathcal{N} = \dim\mathcal{L} - \dim\mathcal{A} - \dim\mathcal{K}$, while $\dim\mathcal{A}$ can be determined in all cases by an argument along the lines indicated in Sec. V.

B. Examples

1. $\mathcal{L} = so(4,1)$

The positive roots of $\tilde{\mathcal{L}} = B_2$ are α_1, α_2, $\alpha_1+\alpha_2$ and $\alpha_1+2\alpha_2$. The chief inner involutive automorphism $Z = \exp(\mathrm{ad}h)$ generating $\mathcal{L} = so(4,1)$ may be chosen[12] so that

$$\exp[\alpha_1(h)] = \exp[(\alpha_1+2\alpha_2)(h)] = 1,$$
$$\exp[\alpha_2(h)] = \exp[(\alpha_1+\alpha_2)(h)] = -1.$$

Thus $\mathcal{K}$ has as its basis ih_α with $\alpha = \alpha_1, \alpha_2$, together with $(e_\alpha + e_{-\alpha})$ and $i(e_\alpha - e_{-\alpha})$ with $\alpha = \alpha_1, \alpha_1+2\alpha_2$. Similarly, $\mathcal{P}$ has basis elements $i(e_\alpha + e_{-\alpha})$ and $(e_\alpha - e_{-\alpha})$ with $\alpha = \alpha_2, \alpha_1+\alpha_2$. Clearly, $\mathcal{A}$ is one-dimensional and may be chosen to have basis element $H_1' = i(e_\alpha + e_{-\alpha})$ with $\alpha = \alpha_2$. $\mathcal{M}$ is also one-dimensional, and its basis element may be taken to be $-iH_2' = ih_\beta$, where $\beta = \alpha_1+\alpha_2$ [as $\alpha_2(h_\beta) = 0$]. Thus $R_{\mathcal{A}} = \{\alpha_2\}$ and $R_{\mathcal{M}}$ is empty, so $V = V_\alpha$ with $\alpha = \alpha_2$. It follows from (16) and (17) that $H_1 = -\{2/(\alpha,\alpha)\}^{1/2}h_\alpha$ with $\alpha = \alpha_2$ and $H_2 = -h_\beta$ with $\beta = \alpha_1+\alpha_2$. On applying criteria (A), (B), and (C) one obtains $P = \{\alpha_1, -\alpha_2, -(\alpha_1+\alpha_2), -(\alpha_1+2\alpha_2)\}$, $P_- = \{-(\alpha_1+\alpha_2)\}$, and $P_+ = \{\alpha_1, -\alpha_2, -(\alpha_1+2\alpha_2)\}$. Thus from Eqs. (A4) and (A8), and as $(\alpha_2,\alpha_2) = \frac{1}{6}$, the basis elements of $\bar{\mathcal{N}}$ are

$$V_{\alpha_2}^{-1}e_{\alpha_1} = \tfrac{1}{2}e_{\alpha_1} + i2^{-1/2}s_1 e_{\alpha_1+\alpha_2} + \tfrac{1}{2}s_2 e_{\alpha_1+2\alpha_2},$$
$$V_{\alpha_2}^{-1}e_{-\alpha_2} = -\tfrac{1}{2}(e_{\alpha_2} - e_{-\alpha_2}) - i3^{1/2}h_{\alpha_2},$$
$$V_{\alpha_2}^{-1}e_{-(\alpha_1+2\alpha_2)} = \tfrac{1}{2}e_{-(\alpha_1+2\alpha_2)} + i2^{-1/2}s_1 s_2 e_{-(\alpha_1+\alpha_2)} + \tfrac{1}{2}s_2 e_{-\alpha_1},$$

where $s_1 = \mathrm{sgn}N_{-\alpha_2,\alpha_1+\alpha_2}$, $s_2 = \mathrm{sgn}(N_{\alpha_2,\alpha_1+\alpha_2}N_{-\alpha_2,\alpha_1+\alpha_2})$. The basis elements of $\mathcal{N}$ may therefore be taken to be

$$\tfrac{1}{2}(e_{\alpha_1} + e_{-\alpha_1}) + i2^{-1/2}s_1(e_{\alpha_1+\alpha_2} + e_{-(\alpha_1+\alpha_2)}) + \tfrac{1}{2}s_2(e_{\alpha_1+2\alpha_2} + e_{-(\alpha_1+2\alpha_2)}),$$
$$\tfrac{1}{2}i(e_{\alpha_1} - e_{-\alpha_1}) - 2^{-1/2}s_1(e_{\alpha_1+\alpha_2} - e_{-(\alpha_1+\alpha_2)}) + \tfrac{1}{2}is_2(e_{\alpha_1+2\alpha_2} - e_{-(\alpha_1+2\alpha_2)}),$$
$$-\tfrac{1}{2}(e_{\alpha_2} - e_{-\alpha_2}) - i3^{1/2}h_{\alpha_2}.$$

2. $\mathcal{L} = so(3,2)$

$\mathcal{L}$ is again a real form of $\tilde{\mathcal{L}} = B_2$. The chief inner involutive automorphism $Z = \exp(\mathrm{ad}h)$ generating $\mathcal{L} = so(3,2)$ may be chosen[12] so that

$$\exp[(\alpha_1+\alpha_2)(h)] = 1,$$
$$\exp[\alpha_1(h)] = \exp[\alpha_2(h)] = \exp[(\alpha_1+2\alpha_2)(h)] = -1.$$

Then $\mathcal{K}$ has basis ih_α with $\alpha = \alpha_1, \alpha_2$ together with $(e_\alpha + e_{-\alpha})$ and $i(e_\alpha - e_{-\alpha})$ with $\alpha = \alpha_1+\alpha_2$, while $\mathcal{P}$ has basis $i(e_\alpha + e_{-\alpha})$ and $(e_\alpha - e_{-\alpha})$ with $\alpha = \alpha_1, \alpha_2$ and $\alpha_1+2\alpha_2$. In this case $m = \dim\mathcal{A} = 2$, and the basis of $\mathcal{A}$ may be chosen so that

$$H_1' = i(e_{\alpha_1} + e_{-\alpha_1}), \quad H_2' = i(e_{\alpha_1+2\alpha_2} + e_{-(\alpha_1+2\alpha_2)}).$$

The dimension of $\mathcal{M}$ is zero, so $R_{\mathcal{A}} = \{\alpha_1, \alpha_1+2\alpha_2\}$, $R_{\mathcal{M}}$ is empty, and

$$V = V_{\alpha_1}V_{\alpha_1+2\alpha_2}.$$

It follows that

$$H_1 = -\{2/(\alpha_1,\alpha_1)\}^{1/2}h_{\alpha_1},$$
$$H_2 = -\{2/(\alpha_1+2\alpha_2, \alpha_1+2\alpha_2)\}^{1/2}\{h_{\alpha_1} + 2h_{\alpha_2}\},$$

where $(\alpha_1,\alpha_1) = (\alpha_1+2\alpha_2, \alpha_1+2\alpha_2) = \frac{1}{3}$. Criteria (A), (B), and (C) give $P = P_+ = \{-\alpha_1, \alpha_2, -(\alpha_1+\alpha_2), -(\alpha_1+2\alpha_2)\}$ with P_- empty. Thus the basis elements of $\bar{\mathcal{N}}$ are [on using Eqs. (A4), (A6), and (A7)]

$$V^{-1}e_{-\alpha_1} = -\tfrac{1}{2}(e_{\alpha_1} - e_{-\alpha_1}) - \tfrac{1}{2}i6^{1/2}h_{\alpha_1},$$
$$V^{-1}e_{\alpha_2} = \tfrac{1}{2}(e_{\alpha_2} - se_{-\alpha_2}) - \tfrac{1}{2}is'(e_{\alpha_1+\alpha_2} - se_{-(\alpha_1+\alpha_2)}),$$
$$V^{-1}e_{-(\alpha_1+\alpha_2)} = \tfrac{1}{2}iss'(e_{\alpha_2} + se_{-\alpha_2}) + \tfrac{1}{2}s(e_{\alpha_1+\alpha_2} + se_{-(\alpha_1+\alpha_2)}),$$
$$V^{-1}e_{\alpha_1+2\alpha_2} = -\tfrac{1}{2}(e_{\alpha_1+2\alpha_2} - e_{-(\alpha_1+2\alpha_2)}) - \tfrac{1}{2}i6^{1/2}\{h_{\alpha_1} + 2h_{\alpha_2}\},$$

where $s = \mathrm{sgn}(N_{\alpha_1,\alpha_2}N_{\alpha_1+2\alpha_2,-\alpha_2})$, $s' = \mathrm{sgn}N_{\alpha_1,\alpha_2}$. The basis elements of $\mathcal{N}$ are then

$$-\tfrac{1}{2}(e_{\alpha_1} - e_{-\alpha_1}) - \tfrac{1}{2}i6^{1/2}h_{\alpha_1},$$
$$\tfrac{1}{2}(e_{\alpha_1+2\alpha_2} - e_{-(\alpha_1+2\alpha_2)}) - \tfrac{1}{2}i6^{1/2}(h_{\alpha_1} + 2h_{\alpha_2}),$$
$$\tfrac{1}{2}(e_{\alpha_2} - e_{-\alpha_2}) - \tfrac{1}{2}is'(e_{\alpha_1+\alpha_2} - e_{-(\alpha_1+\alpha_2)}),$$
$$\tfrac{1}{2}i(e_{\alpha_2} + e_{-\alpha_2}) + \tfrac{1}{2}s'(e_{\alpha_1+\alpha_2} + e_{-(\alpha_1+\alpha_2)}).$$

TABLE V. Dimensions of $\mathcal{L}$, $\mathcal{K}$, $\mathcal{A}$, $\mathcal{M}$, and $\mathcal{N}$ for the simple real forms $\mathcal{L}$ whose complexifications $\tilde{\mathcal{L}}$ are semisimple.

$\mathcal{L}$	$sl(\frac{1}{2}l+1, C)$, l even	$so(l, C)$, l even	$sp(\frac{1}{2}l, C)$, l even	$so(l+1, C)$, l even
$\dim\mathcal{L}$	$\frac{1}{2}l^2+2l$	l^2+l	l^2+l	l^2+l
$\dim\mathcal{K}$	$\frac{1}{4}l^2+l$	$\frac{1}{2}(l^2+l)$	$\frac{1}{2}(l^2+l)$	$\frac{1}{2}(l^2+l)$
$\dim\mathcal{A}$	$\frac{1}{2}l$	$\frac{1}{2}l$	$\frac{1}{2}l$	$\frac{1}{2}l$
$\dim\mathcal{M}$	$\frac{1}{2}l$	$\frac{1}{2}l$	$\frac{1}{2}l$	$\frac{1}{2}l$
$\dim\mathcal{N}$	$\frac{1}{4}l^2+\frac{1}{2}l$	$\frac{1}{2}l^2$	$\frac{1}{2}l^2$	$\frac{1}{2}l^2$

3. $\mathcal{L} = so(4,2)[\sim su(2,2)]$

In this case $\tilde{\mathcal{L}} = D_3(\sim A_3)$, whose positive roots are $\alpha_1, \alpha_2, \alpha_3, \alpha_1+\alpha_2, \alpha_1+\alpha_3, \alpha_1+\alpha_2+\alpha_3$. The chief inner involutive automorphism $Z = \exp(\mathrm{ad}h)$ generating $\mathcal{L} = so(4,2)$ may be chosen[12] so that

$$\exp[\alpha(h)] = \begin{cases} 1, & \alpha = \alpha_2, \alpha_3, \\ -1, & \alpha = \alpha_1, \alpha_1+\alpha_2, \alpha_1+\alpha_3, \alpha_1+\alpha_2+\alpha_3. \end{cases}$$

This implies that $\mathcal{K}$ has basis ih_α with $\alpha = \alpha_1, \alpha_2, \alpha_3$, together with $(e_\alpha + e_{-\alpha})$ and $i(e_\alpha - e_{-\alpha})$ with $\alpha = \alpha_2, \alpha_3$, whereas $\mathcal{P}$ has basis $i(e_\alpha + e_{-\alpha})$ and $(e_\alpha - e_{-\alpha})$ with $\alpha = \alpha_1, \alpha_1+\alpha_2, \alpha_1+\alpha_3, \alpha_1+\alpha_2+\alpha_3$. $\mathcal{A}$ is two-dimensional, and its basis may be chosen so that

$$H_1' = i(e_{\alpha_1} + e_{-\alpha_1}), \quad H_2' = i(e_{\alpha_1+\alpha_2+\alpha_3} + e_{-(\alpha_1+\alpha_2+\alpha_3)}).$$

$\mathcal{H}$ is one-dimensional with

$$-iH_3' = i(h_{\alpha_2} - h_{\alpha_3}).$$

Thus $R_{\mathcal{A}} \cup R_{\mathcal{H}} = \{\alpha_1, \alpha_1+\alpha_2+\alpha_3\}$ so that

$$V = V_{\alpha_1} V_{\alpha_1+\alpha_2+\alpha_3}$$

and hence

$$H_1 = -\{2/(\alpha_1,\alpha_1)\}^{1/2} h_{\alpha_1},$$

$$H_2 = -\{2/(\alpha_1+\alpha_2+\alpha_3, \alpha_1+\alpha_2+\alpha_3)\}^{1/2}(h_{\alpha_1}+h_{\alpha_2}+h_{\alpha_3}),$$

$$H_3 = -h_{\alpha_2} + h_{\alpha_3},$$

where $(\alpha_1,\alpha_1) = (\alpha_1+\alpha_2+\alpha_3, \alpha_1+\alpha_2+\alpha_3) = \frac{1}{4}$. Criteria (A), (B), and (C) then imply that $P = P_+ = \{-\alpha_1, \alpha_2, \alpha_3, -(\alpha_1+\alpha_2), -(\alpha_1+\alpha_3), -(\alpha_1+\alpha_2+\alpha_3)\}$ with P_- empty. By Eqs. (A4), (A6), and (A7), $\tilde{\mathcal{N}}$ spanned by the basis vectors

$$V^{-1}e_{-\alpha_1} = -\tfrac{1}{2}(e_{\alpha_1} - e_{-\alpha_1}) - i2^{1/2}h_{\alpha_1}, \tag{18}$$

$$V^{-1}e_{-(\alpha_1+\alpha_2+\alpha_3)} = -\tfrac{1}{2}(e_{\alpha_1+\alpha_2+\alpha_3} - e_{-(\alpha_1+\alpha_2+\alpha_3)}) - i2^{1/2}(h_{\alpha_1}+h_{\alpha_2}+h_{\alpha_3}), \tag{19}$$

together with

$$V^{-1}e_{\alpha_2} = \tfrac{1}{2}[e_{\alpha_2} - i\{\mathrm{sgn}N_{\alpha_1+\alpha_2+\alpha_3,-(\alpha_1+\alpha_3)}\}e_{-(\alpha_1+\alpha_3)} - i\{\mathrm{sgn}N_{\alpha_1,\alpha_2}\}e_{\alpha_1+\alpha_2} - \{\mathrm{sgn}(N_{\alpha_1,\alpha_2}N_{\alpha_1+\alpha_2+\alpha_3,-\alpha_3}\}e_{-\alpha_3}],$$

$$V^{-1}e_{-(\alpha_1+\alpha_2)} = \tfrac{1}{2}[e_{-(\alpha_1+\alpha_2)} - i\{\mathrm{sgn}N_{\alpha_1+\alpha_2+\alpha_3,-(\alpha_1+\alpha_2)}\}e_{\alpha_3} - i\{\mathrm{sgn}N_{\alpha_1,-(\alpha_1+\alpha_2)}\}e_{-\alpha_2} - \{\mathrm{sgn}(N_{\alpha_1,-(\alpha_1+\alpha_2)}N_{\alpha_1+\alpha_2+\alpha_3,-\alpha_2})\}e_{\alpha_1+\alpha_3}],$$

and $V^{-1}e_{\alpha_3}$, $V^{-1}e_{-(\alpha_1+\alpha_3)}$, which are given by similar expressions with α_2 and α_3 interchanged. By virtue of the identities[21]

$$N_{\alpha_1+\alpha_2+\alpha_3,-(\alpha_1+\alpha_3)} = -N_{\alpha_1+\alpha_2+\alpha_3,-\alpha_2},$$

$$N_{\alpha_1,-(\alpha_1+\alpha_2)} = -N_{\alpha_1,\alpha_2}$$

(and similar identities with α_2 and α_3 interchanged), and

$$\mathrm{sgn}\{N_{\alpha_1+\alpha_2+\alpha_3,-\alpha_3}\} = \mathrm{sgn}\{N_{\alpha_1,\alpha_2}N_{\alpha_1,\alpha_3}N_{\alpha_1+\alpha_2+\alpha_3,-\alpha_2}\},$$

it follows that the six basis elements of $\mathcal{N}$ may be taken to be (18) and (19) together with

$$\tfrac{1}{2}(e_{\alpha_2} + e_{-\alpha_2}) - \tfrac{1}{2}i\{\mathrm{sgn}N_{\alpha_1,\alpha_2}\}(e_{\alpha_1+\alpha_2} + e_{-(\alpha_1+\alpha_2)}),$$

$$\tfrac{1}{2}i(e_{\alpha_2} - e_{-\alpha_2}) + \tfrac{1}{2}\{\mathrm{sgn}N_{\alpha_1,\alpha_2}\}(e_{\alpha_1+\alpha_2} - e_{-(\alpha_1+\alpha_2)}),$$

and two similar expressions with α_2 replaced by α_3.

V. IWASAWA DECOMPOSITIONS FOR REAL LIE ALGEBRAS GENERATED BY CHIEF OUTER INVOLUTIVE AUTOMORPHISMS

A. General considerations

The chief outer involutive automorphism $Z = Z_0 \exp(\mathrm{ad}h)$ of $\mathcal{L}_c$ and $\tilde{\mathcal{L}}$ is such that[10,11,13] for the simple roots $\alpha_1, \ldots, \alpha_l$

$$Z_0 h_{\alpha_j} = \sum_{k=1}^{l} \tau_{jk} h_{\alpha_k}, \quad Z_0 e_{\pm\alpha_j} = \sum_{k=1}^{l} \tau_{jk} e_{\pm\alpha_k},$$

while for each nonsimple root $\alpha = \sum_{j=1}^{l} \kappa_j \alpha_j$

$$Z_0 e_\alpha = \chi_\alpha e_{\alpha''}, \tag{20}$$

where

$$\alpha'' = \sum_{j=1}^{l}\sum_{k=1}^{l} \tau_{jk}\kappa_j \alpha_k$$

and $\chi_\alpha = \pm 1$. The sign of χ_α is such that $\chi_\alpha = +1$ for α simple, $\chi_{\alpha+\beta} = \{N_{\alpha'',\beta''}/N_{\alpha,\beta}\}\chi_\alpha\chi_\beta$, and $\chi_{-\alpha} = \chi_\alpha$. That is, Z_0 is specified by a "rotation" τ in the root space of $\tilde{\mathcal{L}}$ (or alternatively in $\mathcal{H}$), this rotation corresponding to a transformation that leaves invariant the Dynkin diagram of $\tilde{\mathcal{L}}$.

The major difference from the case considered in Sec. IV is that Z is *not* diagonal if the basis of $\mathcal{L}_c$ is taken to consist of ih_α, $\alpha = \alpha_1, \ldots, \alpha_l$, together with $(e_\alpha + e_{-\alpha})$ and $i(e_\alpha - e_{-\alpha})$ for all α. In general the eigenvectors of Z are linear combinations of the ih_α, $\alpha = \alpha_1, \ldots, \alpha_l$, or of the form $(e_\alpha + e_{-\alpha}) \pm (e_{\alpha''} + e_{-\alpha''})$ and $i(e_\alpha - e_{-\alpha}) \pm i(e_{\alpha''} - e_{-\alpha''})$, where e_α and $e_{\alpha''}$ are related by (20). Only when $\alpha'' = \alpha$ do the simple combinations $(e_\alpha + e_{-\alpha})$ and $i(e_\alpha - e_{-\alpha})$ remain eigenvectors of Z. Thus Z has three types of eigenvector in $\mathcal{L}_c$, namely:

type I: linear combinations of the ih_α, $\alpha = \alpha_1, \ldots, \alpha_l$;

type II: $(e_\alpha + e_{-\alpha}) \pm (e_{\alpha''} + e_{-\alpha''})$ and $i(e_\alpha - e_{-\alpha}) \pm i(e_{\alpha''} - e_{-\alpha''})$ (which occur when $\alpha'' \neq \alpha$);

type III: $(e_\alpha + e_{-\alpha})$ and $i(e_\alpha - e_{-\alpha})$ (which occur when $\alpha'' = \alpha$).

It will be shown in the subsequent subsections that in every case $\mathcal{A}$ and $\mathcal{H}$ can be constructed so that their basis vectors are formed from eigenvectors of types I or III alone. This implies that the automorphism V is either the identity or is again of the form (15).

B. Iwasawa decompositions of the real Lie algebras $= sl(l+1, R)$ and $q_{(l+1)/2}$ of $\tilde{\mathcal{L}} = A_l$ for $l \geq 2$

In this case[13] $\tau_{jk} = \delta_{k,l+1-j}$, $j,k = 1, \ldots, l$, which corresponds to the transformation $\alpha_j \to \alpha_{l+1-j}$, $j = 1, \ldots, l$, of the Dynkin diagram of A_l.

Suppose first that l is even. In this case there is only one real form generated from $\mathcal{L}_c = su(l+1)$ by a chief outer involutive automorphism, namely $\mathcal{L} = sl(l+1, R)$, for which $Z = Z_0$ [i. e., $Z = Z_0 \exp(\mathrm{ad}h)$ with $h = 0$]. $\mathcal{P}$ contains the $\frac{1}{2}l$ basis elements

$$h_{\alpha_j} - h_{\alpha_{l+1-j}}, \quad j = 1, 2, \ldots, \tfrac{1}{2}l. \tag{21}$$

The only eigenvectors of Z that commute with the set (21) are those of type III, which occur[13] for the $\frac{1}{2}l$ roots

$$\alpha=\sum_{k=j}^{l/2}(\alpha_{l/2-k+1}+\alpha_{l/2+k}),\quad j=1,2,\ldots,\tfrac{1}{2}l. \tag{22}$$

As $\chi_\alpha=-1$ for all such α, these eigenvectors all give members of $\mathcal{P}$. Moreover, the set of $\frac{1}{2}l$ elements $i(e_\alpha+e_{-\alpha})$, where α is given by (22), mutually commute with each other. Thus $\mathcal{A}$ may be chosen so that its basis consists of this set and the set (21), so that $\dim\mathcal{A}=l$. Hence $\dim\mathcal{H}=0$, so $R_{\mathcal{A}}\cup R_{\mathcal{H}}$ is the set (22) and V is given by (15).

Now suppose that l is odd. Then both $\mathcal{L}=sl(l+1,R)$ and $q_{(l+1)/2}$ are generated from $\mathcal{L}_c=su(l+1)$ by chief outer involutive automorphisms $Z=Z_0\exp(\mathrm{ad}h)$. In both cases $\mathcal{P}$ contains the $\frac{1}{2}(l-1)$ basis elements

$$h_{\alpha_j}-h_{\alpha_{l+1-j}},\quad j=1,2,\ldots,\tfrac{1}{2}(l-1). \tag{23}$$

The only eigenvectors of Z that commute with the set (23) are those of type III, which now occur for the $\frac{1}{2}(l+1)$ roots[13]

$$\alpha=\sum_{k=-j}^{j}\alpha_{k+(l+1)/2},\quad j=0,1,\ldots,\tfrac{1}{2}(l-1). \tag{24}$$

However, for all such α $\chi_\alpha\exp[\alpha(h)]=-1$ for $\mathcal{L}=sl(l+1,R)$ while $\chi_\alpha\exp[\alpha(h)]=+1$ for $\mathcal{L}=q_{(l+1)/2}$, so these eigenvectors give members of $\mathcal{P}$ for $sl(l+1,R)$ but for $q_{(l+1)/2}$ they are members of $\mathcal{K}$. In both cases the set of $\frac{1}{2}(l+1)$ elements $i(e_\alpha+e_{-\alpha})$ commute with each other. Thus, for $\mathcal{L}=sl(l+1,R)$, $\mathcal{A}$ may be chosen so that its basis consists of this set and the set (23), so that $\dim\mathcal{A}=l$, $\dim\mathcal{H}=0$, $R_{\mathcal{A}}\cup R_{\mathcal{H}}$ is the set (24), and V is again given by (15). On the other hand, for $\mathcal{L}=q_{(l+1)/2}$, $\mathcal{A}$ may be chosen so that its basis is the set (23) alone. Then $\dim\mathcal{A}=\frac{1}{2}(l-1)$ and hence $\dim\mathcal{H}=\frac{1}{2}(l+1)$. The simplest choice of basis elements of $\mathcal{H}$ is

$$ih_{\alpha_{(l+1)/2}},ih_{\alpha_1}+ih_{\alpha_l},ih_{\alpha_2}+ih_{\alpha_{l-1}},\ldots,ih_{\alpha_{(l-1)/2}}+ih_{\alpha_{(l+3)/2}}.$$

Thus, for $\mathcal{L}=q_{(l+1)/2}$, $\mathcal{H}'$ and $\mathcal{H}$ coincide, and V can be taken to be the identity mapping.

C. Iwasawa decompositions of the real Lie algebra $=so(2l-2p-1,2p+1)$ $(o\leqslant p<\frac{1}{2}l)$ of $\mathcal{L}=D_l$ for $l\geqslant 2$

In this case[14] $\tau_{jk}=\delta_{jk}$, $j,k=1,2,\ldots,l-2$, and $\tau_{l-1,k}=\delta_{lk}$, $\tau_{lk}=\delta_{l-1,k}$, $k=1,\ldots,l$, which corresponds to interchanging α_{l-1} with α_l in the Dynkin diagram, all the other roots $\alpha_1,\ldots,\alpha_{l-2}$ remaining unchanged. Moreover, $\chi_\alpha=+1$ for all α. It follows from the form of τ that

$$h''=h_{\alpha_{l-1}}-h_{\alpha_l} \tag{25}$$

is a member of $\mathcal{P}$. In order to distinguish the eigenvectors of Z of type III it is convenient to introduce[10,12,14] the l linear functions $\epsilon_j(h)$, $j=1,\ldots,l$, $h\in\mathcal{H}$, such that

$$\epsilon_j(h)=\begin{cases}\alpha_j(h)+\alpha_{j+1}(h)+\cdots+\alpha_{l-2}(h)+\frac{1}{2}\alpha_{l-1}(h)+\frac{1}{2}\alpha_l(h), & j\leq l-2,\\ \frac{1}{2}\alpha_{l-1}(h)+\frac{1}{2}\alpha_l(h), & j=l-1,\\ -\frac{1}{2}\alpha_{l-1}(h)+\frac{1}{2}\alpha_l(h), & j=l.\end{cases}$$

Then $\epsilon_j(h)\pm\epsilon_k(h)$ for $1\leq j<k\leq l$ are the set of positive roots of $\mathcal{L}$ with respect to $\mathcal{H}$. Z_0 corresponds to replacing $\epsilon_l(h)$ by $-\epsilon_l(h)$ with $\epsilon_1(h),\epsilon_2(h),\ldots,\epsilon_{l-1}(h)$ all remaining unchanged. Thus the roots α corresponding to eigenvectors of Z of type III are $\epsilon_j(h)\pm\epsilon_k(h)$ for $1\leq j<k\leq l-1$. These are the only eigenvectors which commute with the element h'' of (25). A mutually commuting subset is given by $(e_\alpha+e_{-\alpha})$ with

$$\alpha=\epsilon_j\pm\epsilon_k,$$
$$(j,k)=\begin{cases}(1,l-1),(2,l-2),\ldots,(\frac{1}{2}(l+1),\frac{1}{2}(l+1)), & l\text{ odd},\\ (1,l-1),(2,l-2),\ldots,(\frac{1}{2}l-1,\frac{1}{2}l+1), & l\text{ even}.\end{cases} \tag{26}$$

For $\mathcal{L}=so(2l-2p-1,2p+1)$ with $p\geq 1$ the chief outer involutive automorphism $Z=Z_0\exp(\mathrm{ad}h)$ generating $\mathcal{L}$ may be chosen[14] so that $\exp[\alpha_{l-p-1}(h)]=-1$ and $\exp[\alpha_j(h)]=+1$ for $j\neq l-p-1$. Then

$$\exp[\epsilon_j(h)]=\begin{cases}1, & l-p-1<j,\\ -1, & j\leq l-p-1.\end{cases} \tag{27}$$

Similarly,[14] for $p=0$, $\exp[\alpha_{l-1}(h)]=\exp[\alpha_l(h)]=-1$ and $\exp[\alpha_j(h)]=1$ for $j=1,\ldots,l-2$, which again implies (27). As $\exp[\epsilon_j(h)\pm\epsilon_k(h)]=-1$ if and only if $j\leq l-p-1<k$, it follows that $\mathcal{A}$ may be chosen to have as its basis the element h'' of (25) together (for $p\geq 1$) with $i(e_\alpha+e_{-\alpha})$ with $\alpha=\epsilon_j\pm\epsilon_k$ and $(j,k)=(1,l-1),(2,l-2),\ldots,(p,l-p)$. Thus $\dim\mathcal{A}=2p+1$ and so $\dim\mathcal{H}=l-2p-1$.

For $p\geq 1$ there are two subcases to be considered. If l is odd then $\mathcal{H}$ may be chosen to have as its basis the $(l-2p-1)$ elements $(e_\alpha+e_{-\alpha})$ with $\alpha=\epsilon_j\pm\epsilon_k$ and $(j,k)=(p+1,l-p-1)$, $(p+2,l-p-2),\ldots,(\frac{1}{2}(l-1),\frac{1}{2}(l+1))$. Then $R_{\mathcal{A}}\cup R_{\mathcal{H}}$ is the set (26), and V is again given by (15). If l is even then $\mathcal{H}$ can be taken to have as its basis the $l-2p-2$ elements $(e_\alpha+e_{-\alpha})$ with $\alpha=\epsilon_j\pm\epsilon_k$ and $(j,k)=(p+1,l-p-1),(p+2,l-p-2),\ldots,(\frac{1}{2}l-1,\frac{1}{2}l+1)$, together with

$$i(h_{\alpha_{l/2}}+h_{\alpha_{l/2+1}}+\cdots+h_{\alpha_{l-2}}+\tfrac{1}{2}h_{\alpha_{l-1}}+\tfrac{1}{2}h_{\alpha_l}).$$

Again $R_{\mathcal{A}}\cup R_{\mathcal{H}}$ is the set (26) and V is given by (15).

For $p=0$, $\dim\mathcal{A}=1$, and $\mathcal{A}$ may be chosen to that h'' of (25) is its basis element. The most convenient choice of base for $\mathcal{H}$ is then

$$ih_{\alpha_1},ih_{\alpha_2},\ldots,ih_{\alpha_{l-2}},ih_{\alpha_{l-1}}+ih_{\alpha_l},$$

so that for $\mathcal{L}=so(2l-1,1)$, $\mathcal{H}'$ and $\mathcal{H}$ coincide, and hence V may be taken to be the identity mapping.

Example. $\mathcal{L}=so(3,1)$ $(\sim sl(2,C))$: In this case $\mathcal{L}=D_2(\sim A_1\oplus A_1)$ whose positive roots are α_1 and α_2. The chief outer automorphism $Z=Z_0\exp(\mathrm{ad}h)$ generating $\mathcal{L}=so(3,1)$ is such that $\exp[\alpha_1(h)]=\exp[\alpha_2(h)]=-1$, so that $\mathcal{K}$ has basis

$$ih_{\alpha_1}+ih_{\alpha_2},\quad\{(e_{\alpha_1}+e_{-\alpha_1})-(e_{\alpha_2}+e_{-\alpha_2})\},$$
$$i\{(e_{\alpha_1}-e_{-\alpha_1})-(e_{\alpha_2}-e_{-\alpha_2})\},$$

and $\mathcal{P}$ has basis

$$h_{\alpha_1}-h_{\alpha_2},\quad i\{(e_{\alpha_1}+e_{-\alpha_1})+(e_{\alpha_2}+e_{-\alpha_2})\},$$
$$\{(e_{\alpha_1}-e_{-\alpha_1})+(e_{\alpha_2}-e_{-\alpha_2})\}.$$

Then $\dim\mathcal{A}=1$, $\dim\mathcal{H}=1$, and the basis elements H_1' of $\mathcal{A}$ and H_2' of $\mathcal{H}$ may be chosen so that

$$H_1' = ih_{\alpha_1} + ih_{\alpha_2}, \quad -iH_2' = h_{\alpha_1} - h_{\alpha_2}.$$

As V can be taken to be the identity mapping $H_1 = H_1'$ and $H_2 = H_2'$. Criteria (A), (B), and (C) imply that $P = P_+ = \{\alpha_1, -\alpha_2\}$, P_- being empty, so that $\tilde{\mathcal{N}}$ has basis $V^{-1}e_\alpha = e_\alpha$ for $\alpha = \alpha_1, -\alpha_2$. The basis elements of $\mathcal{N}$ may therefore be taken to be

$$e_{\alpha_1} - e_{-\alpha_2}, i(e_{\alpha_1} + e_{-\alpha_2}).$$

D. Iwasawa decompositions of the real Lie algebras $= NE_6^3$ and NE_6^4 of $\mathcal{L} = E_6$

The development for this case is very similar to that given in subsection B for the real forms of A_l. The outer automorphism Z_0 may be taken[11,16] to correspond to the transformation of the Dynkin diagram in which $\alpha_1 \longleftrightarrow \alpha_5$, $\alpha_2 \longleftrightarrow \alpha_4$, but α_3 and α_6 are left invariant. ρ therefore contains the two basis elements

$$h_{\alpha_1} - h_{\alpha_5}, \quad h_{\alpha_2} - h_{\alpha_4}. \tag{28}$$

The only eigenvectors of Z that commute with the set (28) are those of type III, which occur[16] for

$$\alpha = \alpha_3, \alpha_6, \alpha_3 + \alpha_6, \alpha_2 + \alpha_3 + \alpha_4, \alpha_2 + \alpha_3 + \alpha_4 + \alpha_6,$$
$$\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5, \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 + \alpha_6,$$
$$\alpha_1 + 2\alpha_3 + \alpha_4 + \alpha_6, \alpha_1 + \alpha_2 + 2\alpha_3 + \alpha_4 + \alpha_5 + \alpha_6,$$
$$\alpha_1 + 2\alpha_2 + 2\alpha_3 + 2\alpha_4 + \alpha_5 + \alpha_6,$$
$$\alpha_1 + 2\alpha_2 + 3\alpha_3 + 2\alpha_4 + \alpha_5 + \alpha_6, \alpha_1 + 2\alpha_2 + 3\alpha_3 + 2\alpha_4 + \alpha_5 + 2\alpha_6,$$

for all of which $\chi_\alpha = +1$.

For $\mathcal{L} = NE_6^4$ all of these type III eigenvectors give members of ρ, so that $\mathcal{A}$ may be chosen to have as its basis the set (28) together with $i(e_\alpha + e_{-\alpha})$ for $\alpha \in R_{\mathcal{A}} = \{\alpha_6, \alpha_2 + 2\alpha_3 + \alpha_4 + \alpha_6, \alpha_1 + \alpha_2 + 2\alpha_3 + \alpha_4 + \alpha_5 + \alpha_6, \alpha_1 + 2\alpha_2 + 2\alpha_3 + 2\alpha_4 + \alpha_5 + \alpha_6\}$. Thus $\dim\mathcal{A} = 6 = l$, so $\dim\mathcal{M} = 0$. Hence $R_{\mathcal{A}} \cup R_{\mathcal{M}} = R_{\mathcal{A}}$ and V is again given by (15).

For $\mathcal{L} = NE_6^3$ all of the type III eigenvectors are members of $\mathcal{K}$, so $\mathcal{A}$ may be chosen to have as its basis

$$ih_{\alpha_1} + ih_{\alpha_5}, ih_{\alpha_2} + ih_{\alpha_4}, ih_{\alpha_3}, ih_{\alpha_6}.$$

As $\mathcal{H}' = \mathcal{H}$, V may be taken to be the identity mapping.

E. Iwasawa decompositions of the simple real Lie algebras $sl(n,C)$, $so(n,C)$, $sp(n,C)$ which are generated from semisimple compact Lie algebras by outer involutive automorphisms

Suppose that $\mathcal{L}_c = \mathcal{L}_{1c} \oplus \mathcal{L}_{2c}$, where $\mathcal{L}_{1c}$ and $\mathcal{L}_{2c}$ are isomorphic simple compact Lie algebras. Let ih_α^k, $\alpha = \alpha_1, \ldots, \alpha_{l/2}$, $(e_\alpha^k + e_{-\alpha}^k)$, $i(e_\alpha^k - e_{-\alpha}^k)$ be the basis of $\mathcal{L}_{kc}$, $k = 1, 2$. (Here l denotes the rank of $\tilde{\mathcal{L}} = \tilde{\mathcal{L}}_1 \oplus \tilde{\mathcal{L}}_2$ and is necessarily even.) For the chief outer involutive automorphism $Z = Z_0 \exp(\mathrm{ad}h)$ generating the simple noncompact real form $\mathcal{L}$ of $\tilde{\mathcal{L}}$ one may take[15] $h = 0$. Z_0 has the property[15] that

$$Z_0 h_\alpha^1 = h_\alpha^2, \quad Z_0 h_\alpha^2 = h_\alpha^1,$$
$$Z_0 e_{\pm\alpha}^1 = e_{\pm\alpha}^2, \quad Z_0 e_{\pm\alpha}^2 = e_{\pm\alpha}^1.$$

Thus $\mathcal{K}$ contains the elements

$$ih_\alpha^1 + ih_\alpha^2, \quad \alpha = \alpha_1, \ldots, \alpha_{l/2}, \tag{29}$$

and ρ contains

$$h_\alpha^1 - h_\alpha^2, \quad \alpha = \alpha_1, \ldots, \alpha_{l/2}. \tag{30}$$

As no other eigenvectors of Z commute with this subset of ρ, $\dim\mathcal{A} = \dim\mathcal{M} = \frac{1}{2}l$, and $\mathcal{A}$ and $\mathcal{M}$ can be taken to have as their bases the sets (30) and (29), respectively. Thus $\mathcal{H}' = \mathcal{H}$ and V can again be taken to be the identity mapping.

APPENDIX: PROPERTIES OF THE AUTOMORPHISM V_α

By definition

$$V_\alpha = \exp\{\mathrm{ad}[ia_\alpha(e_\alpha - e_{-\alpha})]\},$$

where

$$a_\alpha = \pi/[8(\alpha, \alpha)]^{1/2}.$$

Then, as

$$[\exp(\mathrm{ad}A)]B = B + [A, B] + (1/2!)[A, [A, B],] + \cdots, \tag{A1}$$

it follows from Eqs. (1) and (4) that

$$V_\alpha[(e_\alpha + e_{-\alpha})] = i[2/(\alpha, \alpha)]^{1/2}h_\alpha,$$
$$V_\alpha[i(e_\alpha - e_{-\alpha})] = i(e_\alpha - e_{-\alpha}), \tag{A2}$$
$$V_\alpha(ih_\alpha) = -[(\alpha, \alpha)/2]^{1/2}(e_\alpha + e_{-\alpha}).$$

In the application of stage (5) of Sec. III it is necessary to know $V_\alpha^{-1}e_\gamma$ for various roots γ defined with respect to $\mathcal{H}$. There are six cases to be considered:

(1) $\gamma = \alpha$ or $-\alpha$:

$$V_\alpha^{-1}e_\alpha = \tfrac{1}{2}(e_\alpha - e_{-\alpha}) - \tfrac{1}{2}i[2/(\alpha, \alpha)]^{1/2}h_\alpha, \tag{A3}$$
$$V_\alpha^{-1}e_{-\alpha} = -\tfrac{1}{2}(e_\alpha - e_{-\alpha}) - \tfrac{1}{2}i[2/(\alpha, \alpha)]^{1/2}h_\alpha. \tag{A4}$$

(2) The α strong containing γ has only one member and $\gamma \neq \alpha, -\alpha$:

$$V_\alpha^{-1}e_\gamma = e_\gamma. \tag{A5}$$

(3) The α string containing γ has only two members:

(a) if γ is the first member of the string

$$V_\alpha^{-1}e_\gamma = 2^{-1/2}e_\gamma - i2^{-1/2}(\mathrm{sgn}N_{\alpha,\gamma})e_{\gamma+\alpha}; \tag{A6}$$

(b) if γ is the second member of the string

$$V_\alpha^{-1}e_\gamma = -i2^{-1/2}(\mathrm{sgn}N_{\alpha,\gamma-\alpha})e_{\gamma-\alpha} + 2^{-1/2}e_\gamma. \tag{A7}$$

(4) The α string containg γ has only three members:

(a) if γ is the first member of the string

$$V_\alpha^{-1}e_\gamma = \tfrac{1}{2}e_\gamma + i2^{-1/2}(\mathrm{sgn}N_{-\alpha,\gamma+\alpha})e_{\gamma+\alpha} + \tfrac{1}{2}[\mathrm{sgn}(N_{-\alpha,\gamma+\alpha}N_{\alpha,\gamma+\alpha})]e_{\gamma+2\alpha}; \tag{A8}$$

(b) if γ is the second member of the string

$$V_\alpha^{-1}e_\gamma = -i2^{-1/2}(\mathrm{sgn}N_{\alpha,\gamma})e_{\gamma+\alpha} + i2^{-1/2}(\mathrm{sgn}N_{-\alpha,\gamma})e_{\gamma-\alpha}; \tag{A9}$$

(c) if γ is the third member of the string

$$V_\alpha^{-1}e_\gamma = \tfrac{1}{2}e_\gamma - i2^{-1/2}(\mathrm{sgn}N_{\alpha,\gamma-\alpha})e_{\gamma-\alpha} + \tfrac{1}{2}[\mathrm{sgn}(N_{\alpha,\gamma-\alpha}N_{-\alpha,\gamma-\alpha})]e_{\gamma-2\alpha}. \tag{A10}$$

(5) The α string containing γ has four members:

(a) if γ is the first member of the string

$$V_\alpha^{-1}e_\gamma = 8^{-1/2}\{e_\gamma - i3^{1/2}(\mathrm{sgn}N_{\alpha,\gamma})e_{\gamma+\alpha}$$

$$- 3^{1/2}[\mathrm{sgn}(N_{\alpha,\gamma}N_{\alpha,\gamma+\alpha})]e_{\gamma+2\alpha}$$
$$+i[\mathrm{sgn}(N_{\alpha,\gamma}N_{\alpha,\gamma+\alpha}N_{\alpha,\gamma+2\alpha})]e_{\gamma+3\alpha}\}; \quad \text{(A11)}$$

(b) if γ is the second member of the string

$$V_\alpha^{-1}e_\gamma = -8^{-1/2}\{i3^{1/2}(\mathrm{sgn}N_{\alpha,\gamma-\alpha})e_{\gamma-\alpha}+e_\gamma$$
$$+i(\mathrm{sgn}N_{\alpha,\gamma})e_{\gamma+\alpha}+3^{1/2}[\mathrm{sgn}(N_{\alpha,\gamma}N_{\alpha,\gamma+\alpha})]e_{\gamma+2\alpha}\}; \quad \text{(A12)}$$

(c) if γ is the third member of the string

$$V_\alpha^{-1}e_\gamma = -8^{-1/2}\{3^{1/2}[\mathrm{sgn}(N_{\alpha,\gamma-2\alpha}N_{\alpha,\gamma-\alpha})]e_{\gamma-2\alpha}$$
$$+i\,\mathrm{sgn}(N_{\alpha,\gamma-\alpha})e_{\gamma-\alpha}+e_\gamma+i3^{1/2}(\mathrm{sgn}N_{\alpha,\gamma})e_{\gamma+\alpha}\}; \quad \text{(A13)}$$

(d) if γ is the fourth member of the string

$$V_\alpha^{-1}e_\gamma = 8^{-1/2}\{-i[\mathrm{sgn}(N_{\alpha,\gamma-3\alpha}N_{\alpha,\gamma-2\alpha}N_{\alpha,\gamma-\alpha})]e_{\gamma-3\alpha}$$
$$-3^{1/2}[\mathrm{sgn}(N_{\alpha,\gamma-2\alpha}N_{\alpha,\gamma-\alpha})]e_{\gamma-2\alpha}$$
$$+i3^{1/2}(\mathrm{sgn}N_{\alpha,\gamma-\alpha})e_{\gamma-\alpha}+e_\gamma\}. \quad \text{(A14)}$$

Equations (A3) and (A4) follow immediately from (A2), while (A5) is a consequence of the result that $[e_{\pm\alpha}, e_\gamma] = 0$. The remaining equations follow from (A1) by repeated use of (1), (4), and (5), together with the identity $N_{\alpha,\gamma-\alpha} = -N_{\alpha,-\gamma}$ and the fact that if the α string containing γ is $\gamma - r\alpha, \ldots, \gamma, \ldots, \gamma + q\alpha$, then $(N_{\alpha,\gamma})^2 = \frac{1}{2}q(r+1)(\alpha,\alpha)$. These are the only possible cases as the α string containing γ never has more than four members.[21]

Note added in proof: The principal nondegenerate series of unitary irreducible representations of a semisimple noncompact Lie group is induced from a subgroup having $\mathcal{M}\oplus\mathcal{A}\oplus\mathcal{N}$ as its Lie algebra if the set P_- is empty, which is the case for the Lie algebras of all the "complex" groups $sl(n, C)$, $so(n, C)$, and $sp(n, C)$ as well as for many others, including $so(3, 2)$, $so(4, 2)$, and $so(2, 1)$. If P_- is nonempty $\mathcal{M}$ must be replaced by $C(\mathcal{A})$, the centralizer of $\mathcal{A}$ in K, which consists of $\mathcal{M}$ together with the intersection with $\mathcal{L}$ of the subspace of $\mathcal{L}$ generated by $V^{-1}e_\alpha$ and $V^{-1}e_{-\alpha}$ for all $\alpha \in P_-$, and which is therefore easily determined by the methods of this paper.

[1]K. Iwasawa, Ann. Math. **50**, 507 (1949).
[2]G. W. Mackey, *Induced Representations of Groups and Quantum Mechanics* (Benjamin, New York, 1968).
[3]G. W. Mackey, "Group Representations," Lecture Notes, Oxford University, 1967 (unpublished).
[4]F. Bruhat, Bull. Soc. Math. France **84**, 97 (1956).
[5]E. M. Stein, in *High Energy Physics and Elementary Particles*, edited by A. Salam (International Atomic Energy Agency, Vienna, 1965), p. 563.
[6]R. Hermann, *Lie Groups for Physicists* (Benjamin, New York, 1966).
[7]A. Kihlberg, V. F. Müller, and F. Halbwachs, Commun. Math. Phys. **3**, 194 (1966).
[8]M. E. Cartan, Ann. l'Ecole Norm. Sup. 3-ème série, 31, 268 (1914).
[9]M. E. Cartan, J. Math. Pures Appl. **8**, 1 (1929).
[10]F. Gantmacher, Rec. Math. (Mat. Sbornik) N. S. **5**(47), 101 (1939).
[11]F. Gantmacher, Rec. Math. (Mat. Sbornik) N. S. **5**(47), 217 (1939).
[12]J. F. Cornwell, Rep. Math. Phys. **2**, 239 (1971).
[13]J. F. Cornwell, Rep. Math. Phys. **2**, 289 (1971).
[14]J. F. Cornwell, Rep. Math. Phys. **3**, 91 (1972).
[15]J. M. Ekins and J. F. Cornwell, Rep. Math. Phys. **5**, 17 (1974).
[16]J. M. Ekins and J. F. Cornwell, Rep. Math. Phys. (to be published).
[17]J. F. Cornwell and J. M. Ekins, J. Math. Phys. (to be published).
[18]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **15**, 1378 (1974).
[19]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **15**, 1932 (1974).
[20]J. F. Cornwell, Int. J. Theor. Phys. (to be published).
[21]N. Jacobson, *Lie Algebras* (Interscience, New York, 1962), Chap. 4.
[22]Harish-Chandra, Trans. Am. Math. Soc. **75**, 185 (1953), p. 188.

Two-point characteristic function for the Kepler–Coulomb problem

S. M. Blinder

Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48104
(Received 10 March 1975)

Hamilton's two-point characteristic function $S(q_2t_2, q_1t_1)$ designates the extremum value of the action integral between two space–time points. It is thus a solution of the Hamilton–Jacobi equation in two sets of variables which fulfils the interchange condition $S(q_1t_1, q_2t_2) = -S(q_2t_2, q_1t_1)$. Such functions can be used in the construction of quantum-mechanical Green's functions. For the Kepler–Coulomb problem, rotational invariance implies that the characteristic function depends on three configuration variables, say r_1, r_2, r_{12}. The existence of an extra constant of the motion, the Runge–Lenz vector, allows a reduction to two independent variables: $x \equiv r_1 + r_2 + r_{12}$ and $y \equiv r_1 + r_2 - r_{12}$. A further reduction is made possible by virtue of a scale symmetry connected with Kepler's third law. The resulting equations are solved by a double Legendre transformation to yield the Kepler–Coulomb characteristic function in implicit functional form. The periodicity of the characteristic function for elliptical orbits can be applied in a novel derivation of Lambert's theorem.

1. INTRODUCTION

Hamilton' s two-point characteristic function can be defined as the action along a real trajectory connecting two space–time points[1]:

$$S(q_2t_2, q_1t_1) \equiv \int_{t_1}^{t_2} L(q, \dot{q}, t)\, dt. \tag{1}$$

By Hamilton' s principle, the value of the integral between two fixed points represents an extremum wrt variations in path. The function $S(q_2t_2, q_1t_1)$ might not exist for certain pairs of points or might be multivalued for others. The two-point characteristic function is a solution of the Hamilton–Jacobi equation in two sets of variables:

$$\frac{\partial S}{\partial t_2} + H\left(q_2, \frac{\partial S}{\partial q_2}, t_2\right) = 0 \tag{2}$$

and

$$-\frac{\partial S}{\partial t_1} + H\left(q_1, -\frac{\partial S}{\partial q_1}, t_1\right) = 0 \tag{3}$$

The second equation follows from the first by virtue of the interchange condition

$$S(q_1t_1, q_2t_2) = -S(q_2t_2, q_1, t_1) \tag{4}$$

implied by the integral structure of the characteristic function. Initial and final momenta are given by relations of the form

$$p_1 = -\frac{\partial S}{\partial q_1}, \quad p_2 = \frac{\partial S}{\partial q_2}. \tag{5}$$

The two-point characteristic function finds utility in the construction of quantum-mechanical Green' s functions and density matrices.[2] An example is the kernel $K(q_2t_2, q_1t_1)$ which represents a solution of the time-dependent Schrödinger equation

$$\left\{i\hbar\frac{\partial}{\partial t_2} - \mathcal{H}_2\right\} K(q_2t_2, q_1t_1) = 0 \tag{6}$$

subject to the initial condition

$$K(q_2t_1, q_1t_1) = \delta(q_2 - q_1). \tag{7}$$

This Green' s function can be structured in the form

$$K(q_2t_2, q_1t_1) = F(q_2t_2, q_1t_1) \exp\left(\frac{i}{\hbar} S(q_2t_2, q_1t_1)\right) \tag{8}$$

exponentially dependent on the two-point characteristic function. The exchange condition (4) is thus consistent with the Hermitian property

$$K(q_1t_1, q_2t_2)^* = K(q_2t_2, q_1t_1). \tag{9}$$

The preexponential function F in (8) is determined such as to fulfil Eqs. (6) and (7). For the free particle and harmonic oscillator, this is relatively straightforward.

The Coulomb Green' s function $K(\mathbf{r}_2t_2, \mathbf{r}_1t_1)$ has not yet been worked out in closed form,[3] although the time-independent function $G(\mathbf{r}_2, \mathbf{r}_1, E)$ is known.[4] We have attempted to construct the time-dependent function via the representation (8) and have thereby been led to evaluation of the corresponding characteristic function.

2. KEPLER-COULOMB PROBLEM

The Hamilton–Jacobi equation for the attractive Coulomb system reads

$$\frac{\partial S}{\partial t} + \frac{1}{2m}(\nabla S)^2 - \frac{Ze^2}{r} = 0. \tag{10}$$

This pertains as well to the Kepler problem under the substitution $Ze^2 \to GMm$. We are, of course, in the nonrelativistic domain and are assuming $M \gg m$ [or else reading m in Eq. (10) as the reduced mass]. For compactness we shall employ atomic units, setting $m = e = 1$ in Eq. (10). Equivalently, r is to be expressed in units of $a_0 = \hbar^2/me^2$, t in units of $\hbar^3/me^4 = \alpha a_0/c$, and S in units of $\hbar$.

Accordingly, Eqs. (2) and (3) for the Kepler–Coulomb characteristic function take the form

$$\frac{\partial S}{\partial t_2} + \tfrac{1}{2}(\nabla_2 S)^2 - \frac{Z}{r_2} = 0, \tag{11}$$

$$-\frac{\partial S}{\partial t_1} + \tfrac{1}{2}(\nabla_1 S)^2 - \frac{Z}{r_1} = 0.$$

The Hamiltonian is, of course, a constant of the motion, which implies

$$E=-\frac{\partial S}{\partial t_2}=\frac{\partial S}{\partial t_1}. \tag{12}$$

Thus S must depend on t_2 and t_1 only through their difference $t\equiv t_2-t_1$, and

$$E=-\frac{\partial S}{\partial t}. \tag{13}$$

The angular momentum is likewise a constant:

$$\begin{aligned}\mathbf{L}&=\mathbf{r}_1\times\mathbf{p}_1=\mathbf{r}_2\times\mathbf{p}_2\\&=-\mathbf{r}_1\times\nabla_1 S=\mathbf{r}_2\times\nabla_2 S.\end{aligned} \tag{14}$$

Every trajectory is thus confined to the plane normal to the angular momentum vector. One can write

$$\begin{aligned}\nabla_1 S&=\mathbf{u}_1\frac{\partial S}{\partial r_1}+\mathbf{u}_{12}\frac{\partial S}{\partial r_{12}},\\ \nabla_2 S&=\mathbf{u}_2\frac{\partial S}{\partial r_2}+\mathbf{u}_{21}\frac{\partial S}{\partial r_{12}}\end{aligned} \tag{15}$$

in terms of the nonorthogonal unit vectors

$$\begin{aligned}&\mathbf{u}_1\equiv\mathbf{r}_1/r_1,\quad \mathbf{u}_2\equiv\mathbf{r}/r_2,\quad \mathbf{u}_{12}=-\mathbf{u}_{21}\equiv\mathbf{r}_{12}/r_{12},\\ &\mathbf{r}_{12}\equiv\mathbf{r}_1-\mathbf{r}_2,\quad r_{12}\equiv|\mathbf{r}_1-\mathbf{r}_2|.\end{aligned} \tag{16}$$

We find thereby

$$\mathbf{L}=\frac{\mathbf{r}_1\times\mathbf{r}_2}{r_{12}}\frac{\partial S}{\partial r_{12}}=\mathbf{u}_1\times\mathbf{u}_2\frac{r_1 r_2}{r_{12}}\frac{\partial S}{\partial r_{12}}. \tag{17}$$

Thus far, $S(\mathbf{r}_2 t_2, \mathbf{r}_1 t_1)$ has been shown to depend on the four variables r_1, r_2, r_{12}, and t. A further reduction is made possible by the existence of an additional constant of the motion for the Kepler–Coulomb problem, namely the Runge–Lenz vector[5,6]:

$$\mathbf{A}=(Ze^2m)^{-1}\mathbf{L}\times\mathbf{p}+\mathbf{u}. \tag{18}$$

We have therefore

$$\mathbf{A}=Z^{-1}\mathbf{L}\times\nabla_2 S+\mathbf{u}_2=-Z^{-1}\mathbf{L}\times\nabla_1 S+\mathbf{u}_1. \tag{19}$$

The scalar product with $\mathbf{u}_1+\mathbf{u}_2$ results in

$$\mathbf{L}\times(\nabla_1 S+\nabla_2 S)\cdot(\mathbf{u}_1+\mathbf{u}_2)=0. \tag{20}$$

Using (15) and (17), we find thereby

$$\frac{\partial S}{\partial r_1}-\frac{\partial S}{\partial r_2}=0. \tag{21}$$

This shows that S is independent of the variable r_1-r_2; it can depend on r_1 and r_2 only through their sum r_1+r_2. We have thus reduced S to a function of r_1+r_2, r_{12} and t. Cross-derivatives in the Hamilton–Jacobi equation are avoided if one uses as independent variables the linear combinations

$$x\equiv r_1+r_2+r_{12},\quad y\equiv r_1+r_2-r_{12}\quad(0\leq y\leq x<\infty). \tag{22}$$

These are, in fact, the same variables which appear in Lambert's theorem [cf. discussion following Eq. (64)]. The Coulomb Green's function $G(\mathbf{r}_1, \mathbf{r}_2, E)$ was also found to depend on just x and y. Hostler[7] showed that this is likewise a consequence of the "hidden symmetry" associated with the Runge–Lenz vector.

3. SOLUTION OF THE HAMILTON-JACOBI EQUATION

We turn next to the Hamilton–Jacobi equations (11) for the characteristic function $S(x, y, t)$. Using (15) and (22), we find, in terms of the variables x and y,

$$\begin{aligned}\tfrac{1}{2}(\nabla_1 S)^2&=\left(\frac{\partial S}{\partial x}\right)^2+\left(\frac{\partial S}{\partial y}\right)^2+\mathbf{u}_1\cdot\mathbf{u}_{12}\left[\left(\frac{\partial S}{\partial x}\right)^2-\left(\frac{\partial S}{\partial y}\right)^2\right],\\ \tfrac{1}{2}(\nabla_2 S)&=\left(\frac{\partial S}{\partial x}\right)^2+\left(\frac{\partial S}{\partial y}\right)^2+\mathbf{u}_2\cdot\mathbf{u}_{21}\left[\left(\frac{\partial S}{\partial x}\right)^2-\left(\frac{\partial S}{\partial y}\right)^2\right].\end{aligned} \tag{23}$$

Noting that

$$\mathbf{u}_1\cdot\mathbf{u}_{12}-\mathbf{u}_2\cdot\mathbf{u}_{21}=(\mathbf{u}_1+\mathbf{u}_2)\cdot\mathbf{u}_{12}=\frac{r_1-r_2}{r_1 r_2}\frac{xy}{x-y}, \tag{24}$$

the difference between Eqs. (11) reduces to

$$\left(\frac{\partial S}{\partial x}\right)^2-\frac{Z}{x}=\left(\frac{\partial S}{\partial y}\right)^2-\frac{Z}{y}. \tag{25}$$

With the help of (25), the sum of Eqs. (11) works out to

$$\frac{\partial S}{\partial t}+\left(\frac{\partial S}{\partial x}\right)^2+\left(\frac{\partial S}{\partial y}\right)^2-\frac{Z}{x}-\frac{Z}{y}=0. \tag{26}$$

Equations (25) and (26) are equivalent to the symmetrical relations

$$\frac{1}{2}\frac{\partial S}{\partial t}+\left(\frac{\partial S}{\partial x}\right)^2-\frac{Z}{x}=0,\quad \frac{1}{2}\frac{\partial S}{\partial t}+\left(\frac{\partial S}{\partial y}\right)^2-\frac{Z}{y}=0 \tag{27}$$

which have precisely the form of the original Hamilton–Jacobi equations (11) for $L=0$ and r_1, r_2 replaced by $x/2$, $y/2$.

In accordance with Eq. (4), S must fulfil the time-reversal condition

$$S(x, y, -t)=-S(x, y, t) \tag{28}$$

which rules out solutions to (27) obtained simply by separation of variables.

A further symmetry property makes possible a closed-form solution of these coupled equations. This is the invariance of (25)–(27) under the scale transformation: $x, y\to\zeta^2 x, \zeta^2 y$; $t\to\zeta^3 t$; $S\to\zeta S$. Thus

$$S(\zeta^2 x, \zeta^2 y, \zeta^3 t)=\zeta S(x, y, t), \tag{29}$$

showing that S is a linear homogeneous function of the variables $x^{1/2}$, $y^{1/2}$, $t^{1/3}$. The condition (28) is, in fact, a special case of (29), for $\zeta=-1$. By virtue of this homogeneity property, the characteristic function can be represented in the following form: $t^{1/3}\times$ function of $x^{1/2}/t^{1/3}$ and $y^{1/2}/t^{1/3}$.

Specifically, the following definition of variables is convenient:

$$\begin{aligned}&S=(32Z^2t)^{1/3}f(u, v),\\ &u\equiv(x^3/16Zt^2)^{1/6},\quad v\equiv(y^3/16Zt^2)^{1/6}\\ &\text{for } t\geq 0,\quad 0\leq v\leq u<\infty.\end{aligned} \tag{30}$$

Equations (27) thereby transform to

$$\begin{aligned}&\tfrac{4}{3}(f-uf_u-vf_v)+u^{-2}(f_u^2-1)=0,\\ &\tfrac{4}{3}(f-uf_u-vf_v)+v^{-2}(f_v^2-1)=0.\end{aligned} \tag{31}$$

These equations are most readily solved by a double Legendre transformation, whereby

$$F=uf_u+vf_v-f,\quad U=f_u,\quad V=f_v, \tag{32}$$

$$F_U = u, \quad F_V = v.$$

We find thereby

$$F_U = \left(\frac{3}{4}\frac{U^2-1}{F}\right)^{1/2}, \quad F_V = \left(\frac{3}{4}\frac{V^2-1}{F}\right)^{1/2}. \tag{33}$$

The positive square roots are appropriate since $u, v \geq 0$. Some further inequalities are required in order to precisely characterize the solution. Equation (25) implies, since $x \geq y$, that

$$\left|\frac{\partial S}{\partial y}\right| \geq \left|\frac{\partial S}{\partial x}\right|. \tag{34}$$

Since the angular momentum vector is directed parallel to $\mathbf{u}_1 \times \mathbf{u}_2$, Eq. (17) implies that

$$\frac{\partial S}{\partial r_{12}} = \frac{\partial S}{\partial x} - \frac{\partial S}{\partial y} \geq 0. \tag{35}$$

The last two inequalities show that $\partial S/\partial y \leq 0$. Thus, in all cases,

$$f_v \leq 0. \tag{36}$$

For $E \geq 0$, $\partial S/\partial t \leq 0$ and

$$f - uf_u - vf_v \leq 0. \tag{37}$$

Since $f \geq 0$,

$$uf_u + vf_v \geq 0, \quad uf_u - v|f_v| \geq 0. \tag{38}$$

Thus

$$f_u \geq 0 \quad \text{for } E \geq 0. \tag{39}$$

Inequality (38) further implies, in conjunction with (31), that

$$|f_u| \geq |f_v| \geq 1. \tag{40}$$

Combining with (36) and (39),

$$1 \leq f_u < \infty, \quad -\infty < f_v \leq -1. \tag{41}$$

In terms of the transformed variables (32),

$$F \geq 0, \quad 1 \leq U < \infty, \quad -\infty < V \leq -1. \tag{42}$$

It is convenient therefore to define

$$U \equiv \cosh\lambda, \quad V \equiv -\cosh\mu \quad (0 \leq \mu \leq \lambda < \infty). \tag{43}$$

(One might also define a second branch of the function with $0 \geq \mu \geq \lambda > -\infty$ corresponding to points $\mathbf{r}_1, \mathbf{r}_2$ reflected wrt the axis of the Runge–Lenz vector.) Integration of Eq. (33), with the appropriate choice of constant, now gives

$$\frac{8}{3\sqrt{3}}F^{3/2} = \sinh\lambda\cosh\lambda - \lambda - \sinh\mu\cosh\mu + \mu$$
$$= \sinh(\lambda-\mu)\cosh(\lambda+\mu) - (\lambda-\mu). \tag{44}$$

Reversion to the original variables is effected by the inverse transformation:

$$f = UF_U + VF_V - F_U, \quad u = F_U, \quad v = F_V. \tag{45}$$

After some algebra we obtain

$$f(u,v) = u\mathcal{J}(\lambda) - v\mathcal{J}(\mu) \tag{46}$$

where

$$\mathcal{J}(\lambda) \equiv \frac{\sinh\lambda\cosh\lambda + 3\lambda}{4\sinh\lambda} \tag{47}$$

and

$$\frac{\sinh^3\lambda}{u^3} = \frac{\sinh^3\mu}{v^3} = \sinh(\lambda-\mu)\cosh(\lambda+\mu) - (\lambda-\mu). \tag{48}$$

By virtue of (30) and (48), the characteristic function can be expressed in the form

$$S(x,y,t) = (4Zx)^{1/2}\mathcal{J}(\lambda) - (4Zy)^{1/2}\mathcal{J}(\mu). \tag{49}$$

Alternatively,

$$S(\lambda,\mu,t) = \left(\frac{Z^2t}{2}\right)^{1/3}\frac{\sinh(\lambda-\mu)\cosh(\lambda+\mu) + 3(\lambda-\mu)}{[\sinh(\lambda-\mu)\cosh(\lambda+\mu) - (\lambda-\mu)]^{1/3}}. \tag{50}$$

In verification that the preceding represents the solution to Eqs. (25), (26), and (27), it is shown that

$$\frac{1}{2}\frac{\partial S}{\partial t} = -\frac{Z}{x}\sinh^2\lambda = -\frac{Z}{y}\sinh^2\mu \tag{51}$$

$$\frac{\partial S}{\partial x} = \left(\frac{Z}{x}\right)^{1/2}\cosh\lambda, \quad \frac{\partial S}{\partial y} = -\left(\frac{Z}{y}\right)^{1/2}\cosh\mu. \tag{52}$$

Since $\partial S/\partial t = -E$, it follows that $E > 0$ (hyperbolic orbits) is associated with real λ and μ, $E < 0$ (elliptical orbits) with pure imaginary λ and μ. The case $E = 0$ (parabolic orbits) is obtained with $\lambda = \mu = 0$. Equation (48) becomes indeterminate but (49) reduces to

$$S(x,y) = (4Zx)^{1/2} - (4Zy)^{1/2}. \tag{53}$$

This solution does not, however, fulfil the time-reversal condition (28).

When $\mu = 0$, then $v = 0$, $y = 0$ and either $\mathbf{r}_1$ or $\mathbf{r}_2 = 0$. The characteristic function reduces to $S(\mathbf{r}, 0, t)$. As $\lambda \to \mu \neq 0$, $S \to 0$.

The asymptotic region $u, v \to \infty$ pertains to any of the limits $Z \to 0$, $x, y \to \infty$, or $t \to 0$. The asymptotic form of the characteristic function is obtained in the limit $\lambda, \mu \to \infty$, whereby

$$S \sim \left(\frac{Z^2t}{2}\right)^{1/3}\left(\frac{e^{2\lambda} - e^{2\mu}}{4}\right)^{2/3},$$
$$u \sim \frac{e^\lambda}{2}\left(\frac{e^{2\lambda} - e^{2\mu}}{4}\right)^{-1/3}, \quad v \sim \frac{e^\mu}{2}\left(\frac{e^{2\lambda} - e^{2\mu}}{4}\right)^{-1/3}. \tag{54}$$

Thus

$$S \sim \left(\frac{Z^2t}{2}\right)^{1/3}(u^2 - v^2)^2 = \frac{(x-y)^2}{8t} = \frac{r_{12}^2}{2t} \tag{55}$$

which represents the free-particle characteristic function.

4. ELLIPTICAL ORBITS

Negative-energy solutions are most directly obtained by continuation of the variables λ and μ on the imaginary axis. Defining

$$\lambda \equiv i\alpha/2, \quad \mu \equiv i\beta/2 \tag{56}$$

(the factors 1/2 for 2π-periodicity), we obtain

$$S(x,y,t) = (4Zx)^{1/2}F(\alpha) - (4Zy)^{1/2}F(\beta), \tag{57}$$

$$F(\alpha) \equiv \mathcal{J}(i\alpha/2) = \frac{3\alpha + \sin\alpha}{8\sin(\alpha/2)} \tag{58}$$

$$\frac{\sin^3(\alpha/2)}{u^3} = \frac{\sin^3(\beta/2)}{v^3} = \left(\frac{\alpha-\beta}{2}\right) - \sin\left(\frac{\alpha-\beta}{2}\right)\cos\left(\frac{\alpha+\beta}{2}\right). \tag{59}$$

Also, in analogy with (50),

$$S(\alpha,\beta,t) = \left(\frac{Z^2 t/2}{[(\alpha-\beta)/2]-\sin[(\alpha-\beta)/2]\cos[(\alpha+\beta)/2]}\right)^{1/3} \times\left[3\left(\frac{\alpha-\beta}{2}\right)+\sin\left(\frac{\alpha-\beta}{2}\right)\cos\left(\frac{\alpha+\beta}{2}\right)\right]. \quad (60)$$

The characteristic function representing an eliptical orbit should exhibit a periodic structure of the form

$$S(\alpha+n\alpha_0,\beta+n\beta_0,t+n\tau)=S(\alpha,\beta,t)+nS(\alpha_0,\beta_0,\tau), \quad (61)$$
$$n=0,1,2,\cdots,$$

where τ is the period of the orbit. For Eqs. (60) and (61) to be consistent, two conditions must be met:

$$3\left[\left(\frac{\alpha-\beta}{2}\right)+n\left(\frac{\alpha_0-\beta_0}{2}\right)\right]+\sin\left[\left(\frac{\alpha-\beta}{2}\right)+n\left(\frac{\alpha_0-\beta_0}{2}\right)\right] \times\cos\left[\left(\frac{\alpha+\beta}{2}\right)+n\left(\frac{\alpha_0+\beta_0}{2}\right)\right] = 3\left(\frac{\alpha-\beta}{2}\right)+\sin\left(\frac{\alpha-\beta}{2}\right)\cos\left(\frac{\alpha+\beta}{2}\right) +n\left[3\left(\frac{\alpha_0-\beta_0}{2}\right)+\sin\left(\frac{\alpha_0-\beta_0}{2}\right)\cos\left(\frac{\alpha_0+\beta_0}{2}\right)\right] \quad (62)$$

and

$$\frac{t}{[(\alpha-\beta)/2]-\sin[(\alpha-\beta)/2]\cos[(\alpha+\beta)/2]} = \frac{\tau}{[(\alpha_0-\beta_0)/2]-\sin[(\alpha_0-\beta_0)/2]\cos[(\alpha_0-\beta_0)/2]}. \quad (63)$$

The first is most easily fulfilled with $\alpha_0-\beta_0=2\pi$, $\alpha_0+\beta_0=0$. The second gives thereby a relation for the orbital time

$$t=\frac{\tau}{\pi}\left[\left(\frac{\alpha-\beta}{2}\right)-\sin\left(\frac{\alpha-\beta}{2}\right)\cos\left(\frac{\alpha+\beta}{2}\right)\right] =\frac{\tau}{2\pi}[(\alpha-\sin\alpha)-(\beta-\sin\beta)]. \quad (64)$$

This is, in fact, a classical result known as Lambert's theorem.[9] In the original form of the theorem, α and β are defined by

$$\sin\frac{\alpha}{2}\equiv\left(\frac{x}{4a}\right)^{1/2}, \quad \sin\frac{\beta}{2}\equiv\left(\frac{y}{4a}\right)^{1/2}, \quad (65)$$

a being the semimajor axis of the ellipse. By virtue of (51), (13), (56), and the relation $E=-Z/2a$, our definitions of α and β are shown to coincide with (65).

Very similar in form to (64) is Kepler's equation

$$t=\frac{\tau}{2\pi}[(\Theta_2-e\sin\Theta_2)-(\Theta_1-e\sin\Theta_1)] =\frac{\tau}{\pi}\left[\left(\frac{\Theta_2-\Theta_1}{2}\right)-e\sin\left(\frac{\Theta_2-\Theta_1}{2}\right)\cos\left(\frac{\Theta_2+\Theta_1}{2}\right)\right] \quad (66)$$

in which e is the eccentricity and Θ_1, Θ_2 the eccentric anomalies at $\mathbf{r}_1$ and $\mathbf{r}_2$, respectively. Comparing (66) with (64) we can identify

$$\alpha-\beta=\Theta_2-\Theta_1, \quad \cos\left(\frac{\alpha+\beta}{2}\right)=e\cos\left(\frac{\Theta_2+\Theta_1}{2}\right). \quad (67)$$

Setting $\alpha-\beta=2n\pi$, $t=n\tau$ in Eq. (60), we obtain the characteristic function for n complete cycles

$$S=\tfrac{3}{2}n(2\pi Z)^{2/3}\tau^{1/3}. \quad (68)$$

This is related to W, the corresponding solution of the time-independent Hamilton–Jacobi equation, by[10]

$$S=W-Et. \quad (69)$$

Since for elliptical orbits

$$\tau=2\pi Z(-2E)^{-3/2}, \quad (70)$$

we find

$$W=nJ, \quad J=(2\pi Z)^{2/3}\tau^{1/3}, \quad (71)$$

in agreement with the value of the canonical action

$$J=\oint(p_r\,dr+p_\theta\,d\theta+p_\phi\,d\phi). \quad (72)$$

This is equivalent to the more familiar result that

$$E=-2\pi^2Z^2/J^2 \quad (=-2\pi mZ^2e^4/J^2) \quad (73)$$

which for $J=nh$ $(n=1,2,3,\cdots)$ gives the Bohr energy levels.

5. REPULSIVE COULOMB POTENTIAL

For a repulsive Coulomb potential, an analogous calculation leads to the characteristic function

$$S(x,y,t)=(4Zx)^{1/2}\mathcal{G}(\lambda)-(4Zy)^{1/2}\mathcal{G}(\mu), \quad (74)$$

$$\mathcal{G}(\lambda)\equiv\frac{\sinh\lambda\cosh\lambda-3\lambda}{4\cosh\lambda}, \quad (75)$$

$$\frac{\cosh^3\lambda}{u^3}=\frac{\cosh^3\mu}{v^3}=\sinh(\lambda-\mu)\cosh(\lambda+\mu)+(\lambda-\mu) \quad (76)$$

$(0\leq\mu\leq\lambda<\infty)$.

[1] See, for example, J. L. Singe, "Classical Dynamics," in *Handbuch der Physik* Vol. III/1, edited by S. Flügge (Springer, Berlin, 1960), p. 117ff.

[2] R. P. Feynman, Rev. Mod. Phys. **20**, 367 (1948); R. P. Feynman and A.R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965); S.M. Blinder, *Foundations of Quantum Dynamics* (Academic, London, 1974), Chap. 6; S.M. Blinder, "Configuration-Space Green's Functions," in *International Review of Science*, Vol. I, *Theoretical Chemistry* (Butterworths, London, 1975).

[3] For the present status of the problem, see M.J. Goovaerts and J.T. Devreese, J. Math. Phys. **13**, 1070 (1972); R.G. Storer, J. Math. Phys. **9**, 964 (1968).

[4] L. Hostler, J. Math. Phys. **5**, 591 (1964). The two Green's functions are related by Fourier transformation as follows:

$$K(\mathbf{r}_2,\mathbf{r}_1,t)=\lim_{\epsilon\to0}2\pi\int_{-\infty}^{\infty}[G(\mathbf{r}_2,\mathbf{r}_1,E+i\epsilon) -G(\mathbf{r}_2,\mathbf{r}_1,E-i\epsilon)]\,e^{-iEt/\hbar}dE.$$

[5] C. Runge, *Vector Analysis* (Dutton, New York, 1919), p. 79; W. Lenz, Z. Phys. **24**, 197 (1924); W. Pauli, Z. Phys. **36**, 336 (1926) [English translation in B. L. van der Waerden, *Sources of Quantum Mechanics* (Dover, New York, 1968), p. 387]. See also articles by H.V. McIntosh (p. 75) and C.E. Wulfman (p. 145) in *Group Theory and its Applications*, Vol. II, edited by E.M. Loebl (Academic, New York, 1971).

[6] The properties of the Runge-Lenz vector can be developed as follows. Start with Newton's second law for a particle in a Colulomb field:

$$\frac{d\mathbf{p}}{dt} = -\frac{Ze^2}{r^3}\mathbf{r}.$$

Then

$$\mathbf{L}\times\frac{d\mathbf{p}}{dt} = -\frac{Ze^2}{r^3}\mathbf{L}\times\mathbf{r} = -\frac{Ze^2m}{r^3}\left(\mathbf{r}\times\frac{d\mathbf{r}}{dt}\right)\times\mathbf{r}.$$

This works out to

$$\frac{d}{dt}(\mathbf{L}\times\mathbf{p} + Ze^2m\mathbf{u}) = 0,$$

showing that A is a constant of the motion. The equation of the orbit is obtained from

$$\mathbf{A}\cdot\mathbf{r} = Ar\cos\theta = -(Ze^2m)^{-1}L^2 + r,$$
$$r = (Ze^2m)^{-1}L^2/(1 - A\cos\theta),$$

which represents a conic section. The vector A is directed towards the aphelion of the orbit; its magnitude equals the eccentricity.

[7]L. Hostler, J. Math. Phys. **8**, 642 (1967).

[8]This also applies w.r.t. the original position variables:

$$S(\zeta^2\mathbf{r}_1, \zeta^2\mathbf{r}_2, \zeta^3 t) = \zeta S(\mathbf{r}, \mathbf{r}_2, t).$$

Newton's second law for a Coulomb force is likewise invariant under the substitution $\mathbf{r}\to\zeta^2\mathbf{r}$, $t\to\zeta^3 t$. This implies Kepler's third law of planetary motion, that the period of an orbit is proportional to the three-halves power of its linear dimension.

[9]See, for example, E.T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (Cambridge, U.P., Cambridge, 1965), 4th Ed., p. 91-92.

[10]See, for example, H. Goldstein, *Classical Mechanics* (Addison-Wesley, Cambridge, Mass., 1950), p. 299ff.

On the quantum mechanical treatment of dissipative systems

Rainer W. Hasse

*Sektion Physik der Universität München, 8046 Garching, Germany**
and Department of Physics, Indian Institute of Technology Kanpur, Kanpur-208016 (U.P.), India†
(Received 31 October 1974; revised manuscript received 28 March 1975)

Two types of Hamiltonians are investigated which describe quantum mechanically a particle moving subject to a linear viscous force under the influence of a conservative force: the conventional explicitly time-dependent one and an alternative class of nonlinear Hamiltonians. In the latter group we propose a new form. By Ehrenfest's theorem the expectation values of the operators of physical observables correspond to the classical quantities. For all Schrödinger equations we derive and discuss wavepacket, wave, stationary, and pseudostationary solutions of force free motion, free fall, and harmonic oscillator.

1. INTRODUCTION

A. Purpose

Recent heavy ion scattering experiments[1] give strong indication for a new type of reaction called deep inelastic process. The heavy ions lose their entire available kinetic energy during the collision and are then repelled just by their mutual Coulomb interaction energy. Thus nuclear friction seems to play an important role.[2] Moreover, there is also evidence that the fission process is damped during the descent from saddle to scission.[3] Classical[4–6] as well as microscopic[7–9] calculations in these phenomena which include frictional effects have already been made. Quantal friction, however, is still an open problem. Other applications are, for instance, the motion of Stokes' ball in a viscous medium, Brownian motion, or an electric oscillator composed of inductance, capacitor, and resistor.

All examples given above, in principle, can be described by proper many-body Hamiltonians. Their solutions, however, can only be achieved for very simplified systems, e.g., in the model of Ford, Kac, and Mazur,[10] where a particle which is tied to a heat bath consisting of a chain of coupled harmonic oscillators is considered and which then exhibits Brownian motion. Other microscopic models treating radiation damping[11,12] or models of electric circuits or Brownian motion which are based on the interaction of two systems[13–15] shall only be mentioned here. Proceeding this way not only complicates matters but also yields redundant information because the main interest usually is focused only on one particle or, in other words, on a few degrees of freedom. In nuclear physics this fact led to the early introduction of collective variables.[16]

These variables, say X, obey the classical equation of motion $m\ddot{X}+\partial V/\partial X=0$, where $V(X)$ is the potential of a conservative force. Then they are formally quantized by replacing $X\to x$ and $m\dot{x}\to p=-i\hbar\partial/\partial x$, where x,p are now the quantum mechanical variables. The collective system is thus described by a Hamiltonian operator $H(x,p)$ and a wavefunction $\psi(x,t)$, which obeys the Schrödinger equation $i\hbar\partial\psi/\partial t=H\psi$.

It is the purpose of this paper to achieve formal quantizations of a collective "particle" in the above given sense which, in addition to the conservative force $-\partial V/\partial X$, is also subject to a linear frictional force $-\gamma m\dot{X}$. This means physically that some part of the interaction with its surrounding medium or its internal structure is taken into account. No attempt, however, shall be made to calculate the friction coefficient γ microscopically or to apply the formalism to heavy ion scattering or fission.

B. Survey

The formal quantizations can be achieved in two different ways. (i) The historically older method consists in using an explicitly time-dependent Hamiltonian and a canonical momentum which is not associated with the usual momentum. By this circumstance it implies a lack of perspicuity. (ii) The second method introduces a friction potential which contains quantum mechanical expectation values and therefore makes the Schrödinger equation nonlinear. It, hence, has the disadvantage that wavepackets cannot be constructed by superposing waves in the usual fashion.

All Schrödinger equations thus generated are solved for the cases of damped force free motion and free fall and undercritically damped harmonically oscillating motion. In addition to the wavepacket solutions whose centers of gravity travel along the respective classical paths, they also have wave and stationary or pseudostationary solutions.

2. CLASSICAL EQUATION OF MOTION AND CORRESPONDENCE PRINCIPLE

A. Classical equation of motion

The equation of motion of a classical particle moving in one dimension X with momentum $P=m\dot{X}$ through a linearly viscous medium under the influence of a conservative force $-\partial V/\partial X$ reads

$$\dot{P}+\gamma P+\frac{\partial V}{\partial X}=0. \tag{2.1}$$

Then the energy of the particle, i.e., the sum of kinetic $P^2/2m$ and potential V energies,

$$E=T+V, \tag{2.2}$$

is not a constant of motion but decays according to

$$\dot{E}=-(\gamma/m)P^2. \tag{2.3}$$

This energy is dissipated from the particular degree of freedom X, into its surrounding medium or into its internal structure.

B. Correspondence

Quantum mechanically the dissipative system is to be described by a Hamiltonian operator H, which, in its most general form, may depend on the coordinate, the momentum, time, and wavefunction, and a wavefunction $\psi(x,t)$ which is a solution of the time-dependent Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}\psi = H\psi. \tag{2.4}$$

In order to choose a physical description which reproduces the classical results in the limit of large quantum numbers, quantum mechanical correspondence to the classical equations (2.1)–(2.3) has to be established. Let p denote the operator which represents the classical momentum P and let the expectation value $\langle A\rangle$ of an operator A be $\int\psi^* A\psi\,dx$; then for wavepackets which travel along the classical path,

$$\langle x\rangle = X, \quad \langle p\rangle = P, \tag{2.5}$$

we demand Ehrenfest's theorem

$$\left\langle\frac{d}{dt}p\right\rangle + \gamma\langle p\rangle + \left\langle\frac{\partial V}{\partial x}\right\rangle = 0. \tag{2.6}$$

The total time derivative of an operator herein is given by

$$\frac{d}{dt}A = \frac{\partial A}{\partial t} + \frac{i}{\hbar}[H,A]. \tag{2.7}$$

Equation (2.6) is not rigorously equivalent to Eq. (2.1) even for undamped systems because $\langle\partial V/\partial x\rangle = \partial V/\partial X$ is only true for the potentials to be discussed below.[17] In what follows we therefore either have to restrict ourselves to these potentials or take Eq. (2.6) as approximately valid.

Equations (2.2) and (2.3) have no unique quantum mechanical analog since, e.g., $\langle p^2\rangle$, $\langle p\rangle^2$, $\langle p\rangle\langle p^2\rangle^{1/2}$, etc. would all correspond to P^2. Therefore, we can only demand the approximate equalities

$$\langle E\rangle \approx \left\langle\frac{p^2}{2m}\right\rangle + \langle V(x)\rangle, \tag{2.8}$$

$$\left\langle\frac{d}{dt}E\right\rangle \approx -\frac{\gamma}{m}\langle p^2\rangle, \tag{2.9}$$

where we allow for additions of the order of the uncertainty of the respective quantity, e.g.,

$$\Delta_p^2 = \langle p^2\rangle - \langle p\rangle^2. \tag{2.10}$$

Since the energy is not a constant of motion, the definition of the energy operator is still open.

In addition to these Ehrenfest theorems the Heisenberg uncertainty principle

$$\Delta_{\bar p}\,\Delta_x \geq \tfrac{1}{2}\hbar \tag{2.11}$$

is always fulfilled. Here $\bar p$ is the canonical momentum operator which is not necessarily equal to p, and $\Delta_{\bar p}$, Δ_x are given analogously to Eq. (2.10).

As a last point, a solution ψ of Eq. (2.4) should obey a continuity equation which, in its most general form, must be written as

$$\frac{\partial\rho}{\partial t} + \frac{\partial j}{\partial x} = 0,$$

$$\rho = \psi^*\psi, \qquad j = \mathrm{Re}\left(\psi^*\frac{\partial H}{\partial\bar p}\psi\right). \tag{2.12}$$

The generalization of the current is necessary because, according to Hamilton's equation $\dot x = \partial H/\partial\bar p$, the operator corresponding to the velocity is not always given by p/m.

C. Classical solutions

In the following two sections we are studying two classes of Hamiltonians which fulfill the requirements (2.6), (2.8), (2.9), (2.11), (2.12) and achieve their solutions for the problems of force free motion, free fall, and harmonically oscillatory motion. Therefore, we list here the classical solutions of these problems:

Force free motion: $V = 0$

$$X = X_0 + (P_0/m\gamma)(1-\exp(-\gamma t)), \qquad P = P_0\exp(-\gamma t). \tag{2.13}$$

Free fall: $V = -mgX$

$$X = X_0 + (1/\gamma)[gt + (P_0/m - g/\gamma)\times(1-\exp(-\gamma t))],$$

$$P = P_0\exp(-\gamma t) + (mg/\gamma)(1-\exp(-\gamma t)). \tag{2.14}$$

Harmonic oscillator: $V = \frac{1}{2}m\omega^2X^2$

$$X = \exp(-\gamma t/2)\times\left(X_0\cos\Omega t + \frac{P_0/m + \gamma X_0/2}{\Omega}\sin\Omega t\right),$$

$$P = \exp(-\gamma t/2)\times\left(P_0\cos\Omega t - \frac{\gamma P_0/2 + mX_0\omega^2}{\Omega}\sin\Omega t\right) \tag{2.15}$$

Here X_0 and P_0 are the initial values of position and momentum and $\Omega = (\omega^2 - \gamma^2/4)^{1/2}$ is the frequency reduced by the damping. Since we only deal with undercritically damped oscillators, $\gamma < 2\omega$, the reduced frequency is real.

3. EXPLICITLY TIME-DEPENDENT HAMILTONIAN

A. Derivation

There is no direct access to the equation of motion (2.1) via a classical Hamiltonian function $H = T + W$ with the momentum dependent potential $W(X,P)$ because the frictional force $F = -\gamma m\dot X$ cannot be derived from[18] $F = -\partial W/\partial X + (d/dt)(\partial W/\partial\dot X)$. However, the explicitly time-dependent Lagrangian (cf., e.g., Ref. 19)

$$\bar L = \exp(\gamma t)\,[\tfrac{1}{2}m\dot X^2 - V(X)] \tag{3.1}$$

leads to a canonical conjugate momentum

$$\bar P = m\dot X\exp(\gamma t) = P\exp(\gamma t) \tag{3.2}$$

and to the Hamiltonian function

$$H = (\bar P^2/2m)\exp(-\gamma t) + V\exp(\gamma t), \tag{3.3}$$

which gives the correct equation of motion (2.1). Furthermore, the definition of energy

$$E = H\exp(-\gamma t) \tag{3.4}$$

is in accordance with Eqs. (2.2) and (2.3).

By applying the conventional rule of quantization of a classical Hamiltonian the canonical conjugate momentum $\bar{P}$ is replaced by $-i\hbar\partial/\partial x$, thus arriving at the quantum Hamiltonian

$$H = -\frac{\hbar^2}{2m}\exp(-\gamma t)\frac{\partial^2}{\partial x^2} + V(x)\exp(\gamma t) \tag{3.5}$$

and the energy operator $E = H\exp(-\gamma t)$. Equation (3.5) was invented by Kanai,[20] then investigated by several authors,[21-28,37] and has recently been obtained independently.[27,28]

B. Proofs of Ehrenfest theorems

Since $\bar{P} = P\exp(\gamma t) \to \bar{p} = -i\hbar\partial/\partial x$ one finds $p = -i\hbar \times \exp(-\gamma t)\partial/\partial x$ for the operator associated with the classical momentum P. By using the relations $\partial p/\partial t = -\gamma p$ and $[H, p] = [V, \bar{p}] = i\hbar\,\partial V/\partial x$ one readily verifies Eq. (2.6). The expectation value of the energy operator then simply reads $\langle p^2/2m + V\rangle$ and, thus, Eq. (2.8) holds with an equality sign. E commutes with H and therefore $(d/dt)E = \partial E/\partial t = -\gamma p^2/m$, so that (2.9) is fulfilled exactly as well. [Bopp[27] tried to define the energy operator as the Hamiltonian itself and then drew the conclusion that because of $\langle (d/dt)H\rangle = -\gamma\exp(\gamma t)\langle P^2/2m - V\rangle$, Eq. (2.9) is violated.] The continuity equation (2.12) is proved similarly by inserting $\partial H/\partial\bar{p} = \bar{p}\exp(-\gamma t)/m$.

C. Wavepacket solutions

Solutions of the time dependent Schrödinger equation shall be classified as either wavepackets which obey (2.5) or waves, pseudostationary, or stationary solutions, where $\langle x\rangle$ and/or $\langle p\rangle$ either do not exist or are unequal to the classical X, P, respectively.

Historically, Stephens[23] and Kerner[24] were the first to give solutions of Kanai's Hamiltonian (3.5), namely the pseudostationary states of the harmonic oscillator. Other solutions were achieved by Myers[25] (force free motion and free fall waves), Röth[37] (harmonic oscillator wavepacket), Messer[28] (force free motion wavepacket), and by Buch and Denman[26] (force free motion and free fall wavepackets).

It turns out that the wavepacket solutions of all three cases, Eqs. (2.13)–(2.15), can be cast into the same form

$$\psi = \left(\frac{a+a^*}{2\pi a^2}\right)^{1/4}\exp\left(-\frac{(x-X)^2}{2a} + \frac{i}{\hbar}\left[\bar{P}(x-X) + \int\bar{L}\,dt - \theta\right]\right) \tag{3.6}$$

where $\bar{L}$ and $\bar{P}$ are the classical Lagrangian (3.1) and canonical momentum (3.2) and $a(t)$ is a complex and $\theta(t)$ a real function of time different for each case. Its probability density reads

$$\psi^*\psi = \frac{1}{\pi^{1/2}w}\exp\left(-\frac{(x-X)^2}{w^2}\right), \qquad w^2 = \frac{2aa^*}{a+a^*}. \tag{3.7}$$

One easily verifies that (3.6) is a wavepacket in the sense that $\langle x\rangle = X$ and $\langle p\rangle = P$ which obeys the uncertainty equation

$$\Delta_{\bar{p}}\Delta_x = \tfrac{1}{2}\hbar\,|a|/\mathrm{Re}(a). \tag{3.8}$$

However, since $\Delta_p = \Delta_{\bar{p}}\exp(-\gamma t)$, the uncertainty product of position and physical momentum becomes smaller than $\hbar/2$. This unphysical result has already been noted by Brittin[21] and Havas[22] and is explained by Senitzky[15] as due to neglection of a proper fluctuation term induced by the dissipation.

The Fourier transform in the space of the canonical momentum and the canonical momentum probability density read

$$\varphi(\bar{p}) = \left(\frac{a+a^*}{2\pi}\right)^{1/4}\exp\left(-\frac{a}{2\hbar^2}(\bar{p}-\bar{P})^2 - \frac{i}{\hbar}\left[\bar{p}X - \int\bar{L}\,dt + \theta\right]\right), \tag{3.9}$$

$$\varphi^*\varphi = \left(\frac{a+a^*}{2\pi}\right)^{1/2}\exp\left(-\frac{\mathrm{Re}(a)}{\hbar^2}(\bar{p}-\bar{P})^2\right). \tag{3.10}$$

After insertion of the ansatz (3.6) into the Schrödinger equation, one obtains differential equations for a and θ which can be solved:

Force free motion and free fall:

$$\dot{a}(t) = (i\hbar/m)\exp(-\gamma t) \tag{3.11a}$$

$$a(t) = a_0 + (i\hbar/m\gamma)(1-\exp(-\gamma t)) \tag{3.11b}$$

$$w^2(t) = a_0 + (\hbar^2/m^2\gamma^2 a_0)(1-\exp(-\gamma t))^2$$

$$\theta = 0 \tag{3.11c}$$

Harmonic oscillator:

$$a^{-1}(t) = (m/\hbar)(\Omega + \tfrac{1}{2}i\gamma)\exp(\gamma t)$$

$$w^{-2}(t) = (m/\hbar)\Omega\exp(\gamma t), \quad \theta = \tfrac{1}{2}\hbar\Omega t. \tag{3.12}$$

By expanding the exponential in Eq. (3.11), $a(t) \approx a_0 + (i\hbar t/m)(1-\gamma t/2)$, one observes that the widths of the free (force free motion and free fall) wavepackets initially behave like the widths of frictionless wavepackets. For larger times, however, the widths become constant in contrast to the undamped case, cf. the appropriate curves in the figure. The widths of the canonical momentum distributions stay constant for all times.

On the other hand, the width of the harmonic oscillator wavepacket tends to zero, thus contracting to a delta function while its canonical momentum distribution becomes flat. According to the momentum uncertainty $\Delta_p^2 = \hbar^2\exp(-2\gamma t)/(a+a^*)$, however, the momentum distribution contracts to a delta function as well.

D. Wave and pseudostationary solutions

In addition to the wavepackets there also exist other solutions of this Schrödinger equation which do not yield classical results for the expectation values of x and/or p.

For free motion, the wave is easily obtained by observing that the differential equation (3.11a) originally came from $(\dot{a} - i\hbar\exp(-\gamma t)/m)/a^2 = 0$, thus allowing for $1/a = 0$. After inserting this result into (3.6) and rearranging the phase, one obtains

$$\psi_p = \exp\{(i/\hbar)[\bar{P}x - (1/2m)\int P^2\exp(\gamma t)\,dt]\}. \tag{3.13}$$

This plane wave is an eigenfunction of the momentum operator with eigenvalue P, and of the kinetic energy operator. Since $\langle x\rangle$ is entirely undetermined, it is only

an eigenfunction of the energy operator for force free motion but not for free fall. Equation (3.13) can be written as $\exp[(i/\hbar)(\bar{P}x - \int \bar{T}dt)]$, where $\bar{T} = P^2 \exp(\gamma t)/2m$ is the canonical kinetic energy. Therefore, $T = P^2/2m$ can be interpreted as the expectation value of the kinetic energy of this plane wave.

The pseudostationary state solution of the harmonic oscillator is derived similarly,

$$\psi_n = N_n \exp\{[\tfrac{1}{4}\gamma - i\Omega(n+\tfrac{1}{2})]t - (m/2\hbar)(\Omega + \tfrac{1}{2}i\gamma)\exp(\gamma t)x^2\} \times H_n([(m\Omega/\hbar)\exp(\gamma t)]^{1/2}x), \tag{3.14}$$

where $N_n = (m\Omega/\pi\hbar)^{1/4}(2^n n!)^{-1/2}$ and H_n is the Hermite polynomial. The expectation values of x and p both vanish and the wave function (3.14) is neither an eigenfunction of H or E nor is it stationary since $\psi_n^*\psi_n$ depends on time,

$$\psi_n^*\psi_n = N_n^2 \exp[\tfrac{1}{2}\gamma t - (m\Omega/\hbar)\exp(\gamma t)x^2] \times H_n^2([(m\Omega/\hbar)\exp(\gamma t)]^{1/2}x). \tag{3.15}$$

For a given quantum number n, however, the expectation value of the Hamiltonian

$$\langle H\rangle_n = (n+\tfrac{1}{2})\hbar\omega^2/\Omega \tag{3.16}$$

stays constant in time but the expectation value of the energy $\langle E\rangle_n = \langle H\rangle_n \exp(-\gamma t)$ does not.

4. NONLINEAR POTENTIALS

A. Requirements

Another class of dissipative Hamiltonians can be found by circumventing the classical Hamilton's equations. Here one assumes the quantization rule $P \to p = -i\hbar\partial/\partial x$ and searches for a Hermitian potential operator $\gamma W(x, \partial/\partial x; \psi)$ in the Hamiltonian

$$H = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V + \gamma W, \tag{4.1}$$

which gives the expectation value (2.6) of the equation of motion. In addition, since the Hamiltonian (4.1) is not explicitly time dependent, in order to fulfill Eq. (2.8), one identifies the energy operator with the Hamiltonian itself. Then one is led to

$$\langle W\rangle \approx 0, \tag{4.2}$$

$$\left\langle \frac{d}{dt}E\right\rangle = \gamma\left\langle\frac{\partial W}{\partial t}\right\rangle, \tag{4.3}$$

and, by calculating the expectation value of $(d/dt)p = (i/\hbar)[H, p]$, we obtain

$$\left\langle\frac{d}{dt}p\right\rangle + \left\langle\frac{\partial V}{\partial x}\right\rangle + \gamma\left\langle\frac{\partial W}{\partial x}\right\rangle = 0. \tag{4.4}$$

Equation (4.4) together with (2.6) give the defining equation for the additional potential W

$$\frac{\partial W}{\partial x} = \langle p\rangle, \tag{4.5}$$

or, explicitly,

$$\int \psi^* \frac{\partial W}{\partial x}\psi\, dx = -\tfrac{1}{2}i\hbar\int\left(\psi^*\frac{\partial\psi}{\partial x} - \frac{\partial\psi^*}{\partial x}\psi\right)dx. \tag{4.6}$$

B. Kostin's Hamiltonian

One solution of (4.5) due to Kostin[29] has the form

$$\frac{\partial W_K}{\partial x} = -\frac{1}{2}i\hbar\left(\frac{\partial\psi/\partial x}{\psi} - \frac{\partial\psi^*/\partial x}{\psi^*}\right) = -\frac{1}{2}i\hbar\frac{\partial}{\partial x}\ln\frac{\psi}{\psi^*}, \tag{4.7}$$

which in turn when integrated yields

$$W_K = -\tfrac{1}{2}i\hbar\left(\ln\frac{\psi}{\psi^*} - \left\langle\ln\frac{\psi}{\psi^*}\right\rangle\right), \tag{4.8}$$

where the free function of time was chosen according to

$$\langle W_K\rangle = 0. \tag{4.9}$$

The continuity equation can be easily proved by noting that the current is just the ordinary current $j = \mathrm{Re}(\psi^* p\psi)/m$ and $\dot{p} = -(i/\hbar)(\psi^* H\psi - \psi H\psi^*)$. The energy dissipation takes the form

$$\left\langle\frac{d}{dt}H\right\rangle = -\frac{\gamma}{m}\left\langle\left(\frac{\partial W_K}{\partial x}\right)^2\right\rangle = -\gamma m\left\langle\left(\frac{j}{\rho}\right)^2\right\rangle \approx -\frac{\gamma}{m}\langle p^2\rangle. \tag{4.10}$$

The approximate equality in Eq. (4.10) can either be deduced from (4.5) or from the fluid mechanical interpretation[30,31] of quantum mechanics where the current equals the density times the velocity and, hence, mj/ρ is the momentum. As shown by Kan and Griffin,[32] the fluid mechanical interpretation can even serve for deriving Eq. (4.8). One writes $\psi = \varphi\exp(iS/\hbar)$, where φ and S are real functions of position and time and S is the classical action with ∇S equal to the momentum. On the other hand, the phase S of any wavefunction ψ is given by $-\tfrac{1}{2}i\hbar\ln(\psi/\psi^*)$. Thus, by taking the gradient, one arrives at Eq. (4.7) and proves that $\langle\partial W_K/\partial x\rangle$ is equal to the classical momentum.

Solutions of Kostin's nonlinear Hamiltonian are derived in subsections 4D and 4E together with solutions of the following Hamiltonians.

C. Süssmann's Hamiltonian and modifications

Other solutions of Eq. (4.5) for a Hermitian W involve linear combinations of the following operators $\langle x\rangle p$, $\langle p\rangle x$, $\tfrac{1}{2}\langle px + xp\rangle$, $\langle x\rangle\langle p\rangle$, $\tfrac{1}{2}\langle px + xp\rangle$, etc. The only remaining combination however, which simultaneously fulfills Eqs. (4.2) and (2.9) is the general potential[33]

$$W_G = c\{(xp+px)/2 - \langle x\rangle p\} + (1-c)\langle p\rangle(x - \langle x\rangle) = (x - \langle x\rangle)[cp + (1-c)\langle p\rangle] - \tfrac{1}{2}i\hbar c, \tag{4.11}$$

where c may be a real constant or a real function of time. With $\partial H/\partial p = p/m + \gamma c(x - \langle x\rangle)$, the equation of continuity can be proved, and, by differentiating (4.11) with respect to time, one finds

$$\left\langle\frac{d}{dt}H\right\rangle = -\frac{\gamma}{m}\langle p\rangle^2, \tag{4.12}$$

and, by taking the expectation value of Eq. (4.11), one obtains

$$\langle W_G\rangle = c[\langle xp + px\rangle/2 - \langle x\rangle\langle p\rangle] \approx 0. \tag{4.13}$$

The special case of (4.11) with $c = 1$ was found by Süssmann[34] empirically in studying force free motion of wavepackets travelling along classical damped paths. This guess has been inserted into the expression $i\hbar\partial/\partial t$

$+(\hbar^2/2m)\partial^2/\partial x^2 - V(x)$ and the remnants were found to be $W_S\psi$ with

$$W_S = \tfrac{1}{2}[x - \langle x\rangle, p]_+, \tag{4.14}$$

where $[A,B]_+$ is the anticommutator $AB+BA$. As will be shown below, however, the oscillating wavepacket solutions of this Hamiltonian exhibit a peculiar behavior. In order to remove this peculiarity we modified Süssmann's potential and found the two expressions

$$W_{H1} = \tfrac{1}{4}[x - \langle x\rangle,\ p + \langle p\rangle]_+, \tag{4.15}$$

$$W_{H2} = \langle p\rangle(x - \langle x\rangle) - \tfrac{1}{4}[x - \langle x\rangle,\ p - \langle p\rangle]_+, \tag{4.16}$$

which are the special cases $c = \frac{1}{2}$[(4.15)] and $c = -\frac{1}{2}$ [(4.16)] of the general potential (4.11), respectively. Equation (4.16) yet gives unphysical results for the free wavepackets (see next subsection) and therefore has to be abandoned.

Finally, by putting $c=0$ Albrecht[33] obtained

$$W_A = \langle p\rangle(x - \langle x\rangle), \tag{4.17}$$

which gives results very similar to Kostin's Hamiltonian.

D. Wavepacket solutions

Kostin's Hamiltonian (4.8) and the general Hamiltonian (4.11) have the same type of solutions as Kanai's. Since the canonical momentum coincides with the ordinary momentum, the wavepackets read

$$\psi = \left(\frac{a+a^*}{2\pi a^2}\right)^{1/4} \exp\left(-\frac{(x-X)^2}{2a} + \frac{i}{\hbar}\left[P(x-X) + \int L\,dt - \theta\right]\right). \tag{4.18}$$

They obey the uncertainty equation (3.8) with $\bar{p}$ replaced by the momentum variable p and have the corresponding Fourier transforms (3.9), (3.10). For $a(t)$ and $\theta(t)$ we find

	Kostin	General	
Force free motion and free fall :	$\dot{a} + \dfrac{\gamma a(a-a^*)}{2a^*} - \dfrac{i\hbar}{m} = 0$	$\dot{a} - 2c\gamma a - \dfrac{i\hbar}{m} = 0$	(4.19a, b)
	$\theta = -\tfrac{1}{8} i\hbar^2\gamma \int \dfrac{(a^2+a^{*2})(a-a^*)}{aa^*(a+a^*)}\,dt$	$\theta = 0$	(4.20a, b)
Harmonic oscillator:	$a^{-1} = \dfrac{m}{\hbar}\omega$	$a^{-1} = \dfrac{m}{\hbar}[(\omega^2 - c^2\gamma^2)^{1/2} + ic\gamma]$	(4.21a, b)
	$w^{-2} = \dfrac{m}{\hbar}\omega$	$w^{-2} = \dfrac{m}{\hbar}(\omega^2 - c^2\gamma^2)^{1/2}$	(4.22a, b)
	$\theta = \tfrac{1}{2}\hbar\omega t$	$\theta = \tfrac{1}{2}\hbar(\omega^2 - c^2\gamma^2)^{1/2} t.$	(4.23a, b)

Unfortunately, the differential equation (4.19a) for the width $a(t)$ of the force free motion and free fall wavepacket solutions of Kostin's Hamiltonian cannot be solved in closed form. However, the width $w(t)$ as defined in (3.7) obeys the second order differential equation

$$\ddot{w} + \gamma\dot{w} - (\hbar^2/m^2)w^{-3} = 0, \tag{4.24}$$

which, by chance, is just the equation of motion of a classical particle moving under linear friction outside the potential $\hbar^2/2m^2w^2$.

For an undamped free wavepacket one usually employs the initial conditions of a minimum wavepacket,[35] cf. Eq. (3.8), for $t=0$,

$$a(0) = a^*(0) = a_0, \tag{4.25a}$$

$$w(0) = \sqrt{a_0}, \qquad \dot{w}(0) = 0, \tag{4.25b}$$

where (4.25b) follows from (4.25a) but not vice versa. Both sets of initial conditions are also compatible with Kanai's free wavepacket (3.11b, c) and, as can be easily seen, with Kostin's, Eqs. (4.19a), (4.24).

The curve labelled "Kostin" in the figure was calculated by solving Eq. (4.24) numerically with these initial conditions, and its limiting values for small and large times can be obtained in closed form,

$$w^2(t) \to \begin{cases} a_0 + (\hbar^2/m^2a_0)t^2(1-\gamma t) \\ (2\hbar/m)\sqrt{t/\gamma}, \end{cases} \qquad a(t) \to \begin{cases} a_0 + (i\hbar/m)t(1 - \tfrac{1}{2}\gamma t) \\ 4\hbar t/[m(2\sqrt{\gamma t} - i)] \end{cases} \quad \text{for } t \to \begin{cases} 0 \\ \infty. \end{cases} \tag{4.26}$$

For small times Kostin's wavepacket therefore travels like Kanai's. For large times, however, it becomes broader until the probability distribution is entirely flat, cf. the figure. Simultaneously the width of the momentum distribution becomes narrower and finally it goes over into a delta function in such a way that the uncertainty product becomes $\Delta_p\Delta_x = \frac{1}{2}\hbar$. Hence it again becomes a minimum wavepacket contrarily to the frictionless wavepacket whose uncertainty product goes to infinity since the width of its momentum distribution stays finite.

The harmonic oscillator wavepacket solution of Kostin's Hamiltonian was first given by Kan and Griffin.[32] The widths of the probability and momentum distributions stay constant for all times with an uncertainty of exactly $\frac{1}{2}\hbar$. We mention, however, that the

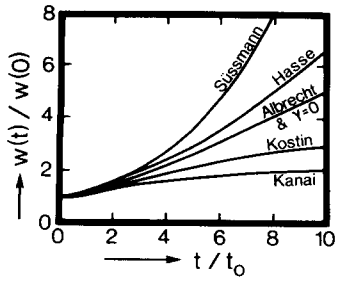


FIG. 1. The widths of damped or undamped ($\gamma = 0$) force free motion and free fall wavepackets calculated with different Hamiltonians. Unit of length $= w(0)$, unit of time $= t_0 = mw^2(0)/2\hbar$, and $\gamma = 1/4t_0$ or 0, initial "velocity" $\dot{w}(0) = 0$.

quantum mechanical frequency in (4.23a) is the original frequency, ω, whereas the classical frequency reduced by damping is $\Omega = (\omega^2 - \gamma^2/4)^{1/2}$. After having lost all available energy, the oscillator ends up in its ground state with energy $\frac{1}{2}\hbar\omega$.

The differential equation (4.19b) for the width of the free wavepackets in the general potential can be solved. In this case, however, the two sets of initial conditions exclude each other. By using (4.25a) one gets

$$w^2 = a_0 \exp(2c\gamma t) + (\hbar^2/m^2c^2\gamma^2 a_0)\sinh^2 c\gamma t,$$
$$a = a_0 \exp(2c\gamma t) + (i\hbar/2mc\gamma)(\exp(2c\gamma t) - 1), \qquad (4.27)$$

and with Eq. (4.25b)

$$w^2 = a_0 \cosh^2 c\gamma t + \frac{\hbar^2}{m^2c^2\gamma^2 a_0}\sinh^2 c\gamma t,$$

$$a = \frac{a_0}{1 + m^2c^2\gamma^2 a_0^2/\hbar^2}\exp(2c\gamma t) + \frac{i\hbar}{2mc\gamma} \times \left(\frac{1 - m^2c^2\gamma^2 a_0^2/\hbar^2}{1 + m^2c^2\gamma^2 a_0^2/\hbar^2}\exp(2c\gamma t) - 1\right), \qquad (4.28)$$

with the difference that (4.27) initially travels like a minimum wavepacket and (4.28) initially travels like a frictionless wavepacket with an initial uncertainty product of $\frac{1}{2}\hbar\,[1 + O(\gamma)]$.

For all values of c the width tends towards infinity as does the solution of Kostin's Hamiltonian. The case $c = 0$ gives exactly the width of an undamped wavepacket and although $a(t)$ is different for $c = +\frac{1}{2}$ and $c = -\frac{1}{2}$, Eq. (4.28) is independent of the sign of c. In the figure the solution (4.28) is plotted.

The widths of the momentum distributions, however, differ. For $c = \frac{1}{2}$ it tends to zero, for $c = 0$ to a constant and for $c = -\frac{1}{2}$ to infinity. An infinite momentum uncertainty, however, means that the expectation value of the kinetic energy becomes infinite according to $\langle p^2\rangle - \langle p\rangle^2 = \Delta_p^2 = \hbar^2/(a + a^*)$ and $T = \langle p^2\rangle/2m$. Since this is an unphysical result, the potential W_{H2} has to be excluded (in free motion there is no energy associated with $\langle x^2\rangle$, and therefore the probability distribution in coordinate space may become infinitely broad).

Süssmann's potential W_S with $c = 1$ gives the fastest increasing width but otherwise similar results as with $c = \frac{1}{2}$. The quantum frequency $\hbar/mw^2$ of the harmonic oscillator solution of the general potential is $(\omega^2 - c^2\gamma^2)^{1/2}$, cf. Eq. (4.21); there is no other difference to the solution of Kostin's Hamiltonian. Albrecht's potential with $c = 0$ even yields the same frequency. The peculiarity mentioned below occurring if one used Süssmann's potential is the frequency $(\omega^2 - \gamma^2)^{1/2}$ which is neither the unshifted frequency ω nor the reduced $\Omega = (\omega^2 - \gamma^2/4)^{1/2}$. Thus, e.g., if $\gamma = \omega$, the classical oscillator undergoes damped oscillations whereas the quantum mechanical wavepacket does not oscillate any more but has a flat probability distribution. The form (4.14) accordingly has to be given up on physical grounds.

In the case $c = \frac{1}{2}$ one obtains the reduced frequency Ω as quantum frequency which seems to be more realistic than ω, which one obtains with the potentials W_K and W_A.

E. Wave and stationary solutions

In his original paper[29] Kostin derived the force free motion wave solution of his Hamiltonian, noting the fact that the expression $\langle \ln(\psi/\psi^*)\rangle$ is undetermined because $\langle x\rangle$ is uncertain. This solution and the corresponding free fall wave and also the free waves in the general potential are very similar to Kanai's,

$$\psi_p = \exp[(i/\hbar)(Px - \int (P^2/2m - \gamma P\langle x\rangle)\,dt) - \tfrac{1}{2}c\gamma t], \qquad (4.29)$$

where $c = 0$ for W_K and W_A. They are eigenfunctions of the momentum and kinetic energy operators, but the total energy is undetermined on grounds of the appearance of $\langle x\rangle$ in the time dependent phase. Furthermore, it is stationary for $c = 0$ but decays if $c > 0$. It would increase exponentially in time if $c < 0$, which again is unphysical.

The solutions of the harmonic oscillator are

$$\psi_n = N_n \exp[-i\omega_c(n + \tfrac{1}{2})t - (m/2\hbar)(\omega_c + ic\gamma)x^2]\,H_n(\sqrt{m\omega c/\hbar}\,x), \qquad (4.30)$$

Kostin, Albrecht: $c = 0, \quad \omega_c = \omega$ (4.31a)

General: $\omega_c = \sqrt{\omega^2 - c^2\gamma^2}$ (4.31b)

Hasse: $c = \pm\frac{1}{2}, \quad \omega_c = \Omega,$ (4.31c)

where $N_n = (m\omega_c/\pi\hbar)^{1/4}(2^n n!)^{-1/2}$ and $\psi^*\psi$ does not depend on time. The solutions (4.30) are therefore stationary. Kostin's potential W_K here vanishes identically since ψ/ψ^* does not depend on the coordinate x and Albrecht's potential W_A vanishes as well since $\langle p\rangle = 0$. Thus (4.30) together with (4.31a) are eigenfunctions of the Hamiltonian $T + V$ and, therefore, also eigenfunctions of the energy operator with eigenvalues $E_n = (n + \frac{1}{2})\hbar\omega$. Contrarily, for $c = \frac{1}{2}$, (4.30) with (4.31c) are only eigenfunctions of H with eigenvalues $(n + \frac{1}{2})\hbar\Omega$.

5. CONCLUSION AND OUTLOOK

Kanai's Hamiltonian and the nonlinear potentials W_K, W_A, W_{H1} all have the common feature that they reproduce the results of undamped quantum mechanical systems for fixed time and the limit $\gamma \to 0$. For $\gamma \neq 0$ the expectation values of observables of damped systems correspond to classical quantities. Other features such as the widths of the free (force free motion and free fall) wavepackets, frequency of quantum oscillation,

and action uncertainties, however, differ widely. The formal quantization of dissipative systems under such weak conditions is therefore not unique. Furthermore, there are no experimental tests for the widths, quantum frequency, or action which could serve as decisive factors for one or the other form of quantization.

The most complete analogy to the undamped case is supplied by Albrecht's potential, where one merely has to insert the classical position and momentum of the damped motion into the wavefunctions of the undamped Hamiltonian. This, however, is only true for the three types of one-dimensional motions studies below. Except for Kostin's Hamiltonian, generalization to more space coordinates is straightforward and further investigations into one or more dimensional types of motion other than the ones discussed above (e.g., Coulomb or square well potentials) may furnish further insight into the differences between the three Hamiltonians. (After submitting this paper the author received a preprint of Immele *et al.*,[36] where, besides the force free motion wavepacket, the cases of two interpenetrating wavepackets and barrier penetration have also been solved with Kostin's Hamiltonian.)

It remains unsatisfactory that Kostin's and Albrecht's potentials give harmonic oscillator solutions which oscillate quantum mechanically with the unshifted frequency instead of the reduced classical damped frequency in contrast to Kanai's Hamiltonian and to Hasse's potential.

It also seems to be peculiar that the width of a free wavepacket calculated with Kanai's explicitly time-dependent Hamiltonian tends to a constant when time proceeds, which is in opposition to all other frictional Hamiltonians and also to the undamped one.

Finally, we mention the paradox[32] of the existence of the stationary states with constant energy expectation values. Together with other findings, this will be discussed separately by Albrecht.[33]

ACKNOWLEDGMENTS

The author gratefully acknowledges discussions with Dr. K. Albrecht and Dr. G. Süssmann on this subject. He is indebted to Dr. W.M. Visscher for communicating to him Ref. 29.

*Permanent address.
†Visiting Professor at Kanpur.

[1]M. Lefort, C. Ngô, J. Péter, and B. Tamain, Nucl. Phys. A **216**, 166 (1973).
[2]R.W. Hasse, Proc. Int. Workshop III on Gross Properties of Nuclei and Nuclear Excitations, Hirschegg, Austria, Jan. 1975, AED-Conf-75-009-000, p. 22
[3]W.J. Swiatecki and S. Bjørnholm, Phys. Rep. **4**, 325 (1972).
[4]R. Wieczorek, R.W. Hasse, and G. Süssman, in *Proc. Third Int. Symp. on the Physics and Chemistry of Fission, Rochester*, 1973 (IAEA, Vienna, 1974), Vol. 1, p. 523.
[5]D.H.E. Gross and H. Kalinowski, Phys. Lett. B **48**, 302 (1974).
[6]J.P. Bondorf, M.I. Sobel and D. Sperber, Phys. Rep. **15**, 84 (1974).
[7]G. Schütte and L. Wilets, in Ref. 4, Vol. I, p. 503. *Physics and Chemistry of Fissior, Rochester* 1974 (IAEA, Vienna 1975), Vol. I, p. 503.
[8]D. Glas and U. Mosel, Phys. Lett. B **49**, 301 (1974).
[9]K.T.R. Davies, S.E. Koonin, J.R. Nix, and A.J. Sierk, Ref. 2, p. 8.
[10]G.W. Ford, M. Kac, and P. Mazur, J. Math. Phys. **6**, 504 (1965).
[11]F. Bloch, Physik Z. **29**, 58 (1928).
[12]H. Bauer and J.H.D. Jensen, Z. Physik **124**, 580 (1948).
[13]H.B. Callen and T.A. Welton, Phys. Rev. **83**, 34 (1951).
[14]J. Weber, Phys. Rev. **90**, 977 (1953); **94**, 211 (1954); **101**, 1619 (1956).
[15]I.R. Senitzky, Phys. Rev. **119**, 670 (1960).
[16]A. Bohr, Kgl. Dan. Vid. Selsk. Mat. Fys. Medd. **26**, No. 14 (1952).
[17]A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1964), Vol. I, p. 218.
[18]H. Goldstein, *Classical Mechanics* (Addision-Wesley, Reading, Mass., 1959).
[19]H.H. Denman, Am. J. Phys. **34**, 1147 (1966).
[20]E. Kanai, Prog. Theor. Phys. **3**, 440 (1948).
[21]W.E. Brittin, Phys. Rev. **77**, 396 (1950).
[22]P. Havas, Bull. Am. Phys. Soc. **1**, 337 (1956); Nuovo Cimento Suppl. **3**, 363 (1957).
[23]W.K.H. Stephens, Proc. Phys. Soc. London **72**, 1027 (1958).
[24]E.H. Kerner, Can. J. Phys. **36**, 371 (1958).
[25]V.W. Myers, Am. J. Phys. **27**, 507 (1959).
[26]L.H. Buch and H.H. Denman, Am. J. Phys. **42**, 304 (1974).
[27]F. Bopp, Z. Angew. Phys. **14**, 699 (1962).
[28]J. Messer, Diplomarbeit Darmstadt (1974), unpublished.
[29]M.D. Kostin, J. Chem. Phys. **57**, 3589 (1972); J. Stat. Phys. **12**, 145 (1975).
[30]E. Madelung, Z. Physik **40**, 322 (1927).
[31]D. Bohm, Phys. Rev. **85**, 166 (1952).
[32]K.-K. Kan and J.J. Griffin, Phys. Lett. B **50**, 241 (1974).
[33]K. Albrecht, Phys. Lett. B **56**, 127 (1975).
[34]G. Süssmann, private communication.
[35]L.I. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1955), p. 58.
[36]J.K. Immele, K.-K. Kan, and J.J. Griffin, Nucl. Phys. A **241**, 47 (1975).
[37]G. Röth, Diplomarbeit Frankfurt (1964), unpublished.

The renormalized projection operator technique for nonlinear stochastic equations. III

Vincent A. LoDato

Marist College Research Institute, Poughkeepsie, New York 12601
(Received 11 September 1974)

The solution of the nonlinear stochastic equation $L(x,t,\omega)\,\psi(x,t,\omega) = g(x,t)+f[\psi(x,t,\omega)]$ is found via the renormalized projection operator technique and is approximated to be

$$\langle\psi(x,t)\rangle = \int\int dx'\,dt'\,\langle G_p(x,t|x',t')\rangle\,\{\psi_{\rm int}(x',0)+g(x',t')\} - \int\int dx'\,dt'\langle P(x',t')\rangle\,\{\int\int dx''\,dt''$$
$$\times\langle G_p(x,t|x',t')\;G_0(x',t'|x'',t'')\rangle\,[\psi_{\rm int}(x'',0)+g(x'',t'')]\}+\int\int dx'\,dt'\langle G_p(x,\,t|x',t'\rangle\,\langle f[\psi^H(x',t')]\rangle.$$

The terms $\langle G_p(x,t|x',t')\rangle$ and $\langle G_p(x,t|x',t')G_0\,(x',t'|x'',t'')\rangle$ are the stochastic one- and two-point Green's functions. Also three conditions are shown that the projection operator must have in order to insure convergence.

INTRODUCTION

The area of stochastic equations and in particular nonlinear stochastic equations appear in many representations of physical reality. Problems in plasma turbulence, chemical kinetics in moderately strong turbulent flames, conduction of mass and heat in a turbulent velocity field, birth and death processes, and scattering phenomena in random media exhibit problems of secularity and lack of convergence of the series expansion. The nonlinear case further aggravates the situation by having no suitable way to treat these terms.

In this paper we study a certain class of nonlinear stochastic equations that are of the type $L(x,t,\omega)\psi(x,t,\omega) = g(x,t)+f[\psi]$ and by using the renormalized projection operator[1] (RNPO) technique solve this equation and discuss three facets: (1) the solution of the stochastic Green function with the projection operator $\langle P(x,t)\rangle$ imbedded in $L(x,t,\omega)$, (2) what are the properties of $\langle P(x,t)\rangle$, and finally (3) how does one evaluate the two-point cross correlation function.

The value $\langle\psi(x,t)\rangle$ is defined to be $\int_\Omega\psi(x,t,\omega)P(\omega)\,d\omega$, where $P(\omega)$ is the probability density with the normalization condition $\int_\Omega P(\omega)\,d\omega = 1$. Since the solution of stochastic equations involves $P(\omega)$ which is not known, but usually the expected values of the observables are known along with their cross correlations, we find that the solutions of stochastic equations have all the problems of its counterpart—the deterministic equations; however, we have the added problem of the manner of treating the averaging process.

1. THE NONLINEAR EQUATION

A general equation which describes a host of physical phenomena can be represented as

$$L(x,t,\omega)\psi(x,t,\omega) = g(x,t)+f[\psi(x,t,\omega)], \tag{1.1}$$

where $L(x,t,\omega)$ is a linear random operator with deterministic variables (x,t) and the random variable ω; $g(x,t)$ are the source and/or sink terms associated with the process, and finally $f[\psi(x,t,\omega)]$ is a general nonlinear function of $\psi(x,t,\omega)$. If $\psi(x,t,\omega)$ were concentrations, then $f[\psi]$ might be the nonlinear kinetic processes associated with the creation and annihilation of the species $\psi_i(x,t,\omega)$. However, if ψ describes the heat transport, then $f[\psi]$ could be the nonlinearities associated with the diffusive thermal conductivity.

The integral form of Eq. (1.1) is

$$\psi(x,t,\omega) = \psi_0(x,t,\omega) + L^{-1}(x,t,\omega)g(x,t) + L^{-1}(x,t,\omega)f[\psi(x,t,\omega)], \tag{1.2}$$

where $L(x,t,\omega)\psi_0(x,t,\omega) = 0$ and $\psi^H(x,t,\omega) = \psi_0(x,t,\omega) + L^{-1}(x,t,\omega)g(x,t)$. The random function $\psi^H(x,t,\omega)$ is the solution of the linear part of Eq. (1.1)

Let us now approximate $f[\psi(x,t,\omega)]$ to be

$$f[\psi(x,t,\omega)] = f\{\psi^H(x,t,\omega) + L^{-1}(x,t,\omega)f[\psi(x,t,\omega)]\} \tag{1.3a}$$
$$\approx f[\psi^H(x,t,\omega)] + \langle P(x,t)\rangle L^{-1}(x,t,\omega)f[\psi(x,t,\omega)]. \tag{1.3b}$$

The averaged operator $\langle P(x,t)\rangle$ is a projection operation. Obviously, when $\langle P(x,t)\rangle \approx 0$, then $f[\psi(x,t,\omega)] - f[\psi^H(x,t,\omega)]$ and the projection $\langle P(x,t)\rangle$ physically has the meaning of measuring the deviation of $f[\psi(x,t,\omega)]$ from $f[\psi^H(x,t,\omega)]$. If the nonlinear random function varies linearly, then the operator $\langle P(x,t)\rangle$ may be associated with the gradient of $f[\psi]$ evaluated at $\langle\psi^H(x,t)\rangle$; and in the case when $f[\psi]$ is a vector it could be the Jacobian matrix of $f[\psi]$.

Now, substituting (1.3b) into (1.2), one arrives at

$$\psi(x,t,\omega) = \psi^H(x,t,\omega) + L^{-1}(x,t,\omega)f[\psi^H(x,t,\omega)] + L^{-1}(x,t,\omega)\,\langle P(x,t)\rangle L^{-1}(x,t,\omega)f[\psi]. \tag{1.4}$$

If one iterates (1.4) by using (1.3b) the series becomes

$$\psi(x,t,\omega) = \psi^H(x,t,\omega) + \{1 + L^{-1}(x,t,\omega)\langle P(x,t)\rangle + L^{-1}(x,t,\omega)\langle P(x,t)\rangle L^{-1}(x,t,\omega)\langle P(x,t)\rangle + \cdots\}L^{-1}(x,t,\omega)f[\psi^H(x,t,\omega)] \tag{1.5}$$

or

$$\psi(x,t,\omega) = \psi^H(x,t,\omega) + \{1 - L^{-1}(x,t,\omega)\langle P(x,t)\rangle\}^{-1} \times L^{-1}(x,t,\omega)f[\psi^H(x,t,\omega)] \tag{1.6}$$

and

$$[L(x,t,\omega) - \langle P(x,t)\rangle]\psi(x,t,\omega) = g(x,t) + f[\psi^H(x,t,\omega)] - \langle P(x,t)\rangle\psi^H(x,t,\omega). \tag{1.7}$$

If we just replace

$$f[\psi(x,t,\omega)] - f[\psi^H(x,t,\omega)] + \langle P(x,t)\rangle \times L^{-1}(x,t,\omega)\psi^H(x,t,\omega),$$

then Eq. (1.7) has two distinct features relative to Eq. (1.1). They are: First, the operator $L(x,t,\omega)$ is replaced by $L(x,t,\omega)-\langle P(x,t)\rangle$, i.e., a part of the nonlinear operator is imbedded in the unperturbed operator; physically this is when the kinetic processes are coupled to the transport mechanisms. Secondly, $f[\psi(x,t,\omega)]$ is replaced by $f[\psi^H(x,t,\omega)]-\langle P(x,t)\rangle\psi^H(x,t,\omega)$.

Let us now consider the Green's equation

$$\{L(x,t,\omega)-\langle P(x,t)\rangle\}G_p(x,t|x',t',\omega)=I\delta(x-x')\delta(t-t') \tag{1.8}$$

The subscript p refers to the projection associated with the particular Green's function.

The solution of (1.7) is

$$\psi(x,t,\omega)=\int\int dx'\,dt'G_p(x,t|x',t',\omega)\{\psi_{\rm int}(x',0)+g(x',t') +f[\psi^H(x',t',\omega)]-\langle P(x',t')\rangle\psi^H(x',t',\omega)\}, \tag{1.9}$$

where $\psi_{\rm int}(x',0)$ is an initial deterministic condition. Performing the averaging of (1.9), we find

$$\langle\psi(x,t)\rangle=\int\int dx'\,dt'\langle G_p(x,t|x',t')\rangle\{\psi_{\rm int}(x',0)+g(x',t')\} +\int\int dx'\,dt'[\langle G_p(x,t|x',t')f[\psi^H(x',t')]\rangle -\langle P(x',t')\rangle\langle G_p(x,t|x',t')\psi^H(x',t')\rangle]. \tag{1.10}$$

The problem now is threefold: (1) the solution of the stochastic Green's function, (2) what is the operator $\langle P(x,t)\rangle$, and (3) how do you evaluate the cross correlations in integrand of (1.10). Let us now address ourselves to these three problems.

2. THE SOLUTION OF THE STOCHASTIC GREEN'S FUNCTION

Recently there has been some work on the solution of this class of equations.[1,2] A particular method to this approach is the renormalized projection operator method (the first paper in this series). We shall now apply this approach to the Green's function equation

$$[L(x,t,\omega)-\langle P(x,t)\rangle]G_p(x,t|x',t',\omega)=I\delta(x-x')\delta(t-t'). \tag{2.1}$$

The operator $L(x,t,\omega)$ can be decomposed into a deterministic and stochastic operator assuming $\langle L_1(x,t)\rangle=0$, namely,

$$L(x,t,\omega)\equiv L_0(x,t)+L_1(x,t,\omega). \tag{2.2}$$

Let us now write $G_p(x,t|x',t',\omega)$ as

$$G_p(x,t|x',t',\omega)=\{1+[L_0-\langle P\rangle-\langle L_1(L_0-\langle P\rangle)^{-1}L_1\rangle]^{-1}\langle P_0\rangle\}\times\langle\underset{\sim}{G}(x,t|x',t')\rangle+\delta G(x,t|x',t'\omega), \tag{2.3}$$

where $\delta G(x,t|x',t',\omega)$ is the random fluctuation in the propagator field. The expected value of the field is

$$\langle G_p(x,t)x',t')\rangle=[1+(L_0-\langle P\rangle-\langle L_1(L_0-\langle P\rangle)^{-1}L_1\rangle)^{-1}\langle P_0\rangle]\langle\underset{\sim}{G}\rangle$$

and the projection operator $\langle P_0\rangle$ is any subset of strongly connected graphs.[1]

If we substitute (2.3) into (2.1) and average, we obtain

$$[L_0-\langle P\rangle][1-(L_0-\langle P\rangle-\langle L_1(L_0-\langle P\rangle)^{-1}L_1\rangle)^{-1}\langle P_0\rangle\times\langle\underset{\sim}{G}(x,t|x',t')\rangle =-\langle L_1(x,t)\delta G(x,t|x',t')\rangle+I\delta(x-x')\delta(t-t'). \tag{2.4}$$

If we substract (2.4) from (2.1) and simplify, one arrives at

$$\delta G(x,t|x',t',\omega)=(L_0-\langle P\rangle)^{-1}(L_1\delta G(x,t|x',t',\omega) -\langle L_1(x,t)\delta G(x,t|x',t')\rangle)-(L_0-\langle P\rangle)^{-1}\times L_1[1-(L_0-\langle P\rangle-\langle L_1(L_0-\langle P\rangle)^{-1}L_1\rangle)^{-1}\times\langle P_0\rangle\langle\underset{\sim}{G}(x,t|x',t')\rangle. \tag{2.5}$$

As we defined previously,[1] consider the operator $\hat{H}$

$$\hat{H}=H-\langle H\rangle$$

and the operator

$$\Omega(x,t)G_p(x,t|x',t',\omega)\equiv\widehat{(L_0-\langle P\rangle)^{-1}L_1G_p(x,t|x',t',\omega)} \equiv(L_0-\langle P\rangle)^{-1}(L_1G_p(x,t|x',t',\omega) -\langle L_1(x,t)G_p(x,t|x',t')\rangle).$$

Consequently, we have

$$\Omega(x,t)\langle G_p(x,t|x',t')\rangle=(L_0-\langle P\rangle)^{-1}L_1(x,t)[1-(L_0-\langle P\rangle-\langle L_1(L_0-P)^{-1}L_1\rangle)^{-1}\times\langle P_0\rangle]\langle\underset{\sim}{G}(x,t|x',t')\rangle \tag{2.6}$$

and

$$\Omega(x,t)\delta G(x,t|x',t',\omega)=(L_0-\langle P\rangle)^{-1}(L_1(x,t)\delta G(x,t|x',t',\omega) -\langle L_1(x,t)\delta G(x,t|x',t')\rangle). \tag{2.7}$$

The reason why we defined the operator $\Omega(x,t)$ is because of its property: $\langle\Omega^N(x,t)G_p(x,t|x',t')\rangle=0$ for any function of $G_p(x,t|x',t',\omega)$ and for all n.

From the definition of Ω and (2.5) one arrives at

$$\langle L_1(x,t)\delta G(x,t|x',t')\rangle=\langle L_1(x,t)\Omega(x,t)\delta G(x,t|x',t')\rangle-\langle L_1(x,t)\Omega(x,t)\rangle +[1-(L_0-\langle P\rangle-\langle L_1(L_0-\langle P\rangle)^{-1}L_1\rangle)^{-1}L_1\rangle)^{-1}\langle P_0\rangle]\times\langle\underset{\sim}{G}(x,t|x',t')\rangle \tag{2.8}$$

and thus

$$[L_0-\langle P\rangle-\langle L_1(x,t)\Omega(x,t)\rangle]\times[1-(L_0-\langle P\rangle-\langle L_1(L_0-\langle P\rangle)^{-1}L_1\rangle)^{-1}\langle P_0\rangle] =L_0-\langle P\rangle-\langle L_1(L_0-\langle P\rangle)^{-1}L_1\rangle-\langle P_0\rangle. \tag{2.9}$$

The solution for the mean and fluctuating components become

$$\langle\underset{\sim}{G}(x,t|x',t')\rangle=(L_0-\langle P\rangle-\langle P_0\rangle-\langle L_1(L_0-\langle P\rangle)^{-1}L_1\rangle)^{-1}\times[I\delta(x-x')\delta(t't')+\langle L_1(x,t)\Omega(x,t)\delta G(x,t|x',t')\rangle \tag{2.10a}$$

and

$$\delta G(x,t|x',t',\omega) = -\Omega(x,t)\delta G(x,t|x',t',\omega) - \Omega(x,t)\langle \underset{\sim}{G}(x,t|x',t')\rangle - \Omega(x,t)(L_0(x,t) - \langle P(x,t)\rangle - \langle L_1(x,t) \times [L_0(x,t) - \langle P(x,t)\rangle]^{-1} L_1(x,t))^{-1}\langle P_0(x,t)\rangle \times \langle \underset{\sim}{G}(x,t|x',t')\rangle. \quad (2.10b)$$

Equations (2.10a) and (2.10b) can be solved consecutively, and one concludes that the mean value for the field is

$$\langle G_p(x,t|x',t')\rangle = [L_0(x,t) - \langle P(x,t)\rangle - \langle L_1(x,t)[L_0(x,t) - \langle P(x,t)\rangle]^{-1}L_1(x,t)\rangle - \langle P_0(x,t)\rangle]^{-1} \times \{1 - \langle L_1(x,t)\Omega(x,t)\rangle[1 + [L_0(x,t) - \langle P(x,t)\rangle - \langle L_1(x,t)[L_0(x,t) - \langle P(x,t)\rangle]^{-1}L_1(x,t)\rangle]^{-1} - \langle P_0(x,t)\rangle]^{-1}[L_0(x,t) - \langle P(x,t)\rangle - \langle L_1(x,t)[L_0(x,t) - \langle P(x,t)\rangle]^{-1}L_1(x,t) - \langle P_0(x,t)\rangle]^{-1} + \cdots\} I\delta(x-x')\delta(t-t'). \quad (2.11)$$

By the RNPO method the perturbation series as expressed in terms of powers of

$$[L_0(x,t) - \langle P(x,t)\rangle - \langle P_0(x,t)\rangle - \langle L_1(x,t)[L_0(x,t) - \langle P(x,t)\rangle]^{-1}L_1(x,t)\rangle]^{-1},$$

which is itself a number of infinite series

$$[L_0(x,t) - \langle P(x,t)\rangle - \langle P_0(x,t)\rangle - \langle L_1(x,t)[L_0(x,t) - \langle P(x,t)\rangle]^{-1}L_1(x,t)\rangle]^{-1} = \sum_{k=0}^{\infty}\Bigg[\sum_{k'=0}^{\infty}\Bigg(\sum_{k''=0}^{\infty}\{L_0(x,t)\langle L_1(x,t)[L_0(x,t) - \langle P(x,t)\rangle]^{-1}L_1(x,t)\rangle\}^{k''}L_0^{-1}(x,t)\langle P(x,t)\rangle\Bigg)^{k'} \times \sum_{k'''=0}^{\infty}\{L_0(x,t)\langle L_1(x,t)[L_0(x,t) - \langle P(x,t)\rangle]^{-1}L_1(x,t)\rangle\}^{k'''}L_0^{-1}(x,t)\langle P_0(x,t)\rangle\Bigg]^{k} \times \sum_{k'=0}^{\infty}\Bigg(\sum_{k''=0}^{\infty}\{L_0(x,t)\langle L_1(x,t)[L_0(x,t) - \langle P(x,t)\rangle]^{-1}L_1(x,t)\rangle\}^{k''}L_0^{-1}(x,t)\langle P(x,t)\rangle\Bigg)^{k'} \times \sum_{k'''=0}^{\infty}\{L_0(x,t)\langle L_1(x,t)[L_0(x,t) - \langle P(x,t)\rangle]^{-1}L_1(x,t)\rangle\}^{k'''}L_0^{-1}(x,t)\Bigg]. \quad (2.12)$$

The use of this propagator $[L_0(x,t) - \langle P(x,t)\rangle]^{-1}$ has an infinite number of terms associated with the nonlinearity $f[\psi(x,t,\omega)]$ imbedded in the series. In the case when $\langle P(x,t)\rangle$ is zero, i.e., $f[\psi(x,t,\omega)] = f[\psi^H(x,t,\omega)]$, the propagator becomes

$$[L_0(x,t) - \langle L_1(x,t)L_0^{-1}(x,t)L_1(x,t)\rangle - \langle P_0(x,t)\rangle]^{-1} = \sum_{k=0}^{\infty}\Bigg(\sum_{k'=0}^{\infty}\{L_0(x,t)\langle L_1(x,t)L_0^{-1}(x,t)L_1(x,t)\rangle\}^{k'} \times L_0^{-1}(x,t)\langle P_0(x,t)\rangle\Bigg)^{k}\sum_{k'=0}^{\infty}[L_0^{-1}(x,t) \times \langle L_1(x,t)L_0^{-1}(x,t)L_1(x,t)\rangle]^{k'}L_0^{-1}(x,t). \quad (2.13)$$

The choice of the operator $\langle P_0(x,t)\rangle$ is such that it ensures the convergence of the series (2.12) and if

(1) $\|(L_0 - \langle P\rangle - \langle L_1[L_0 - \langle P\rangle]^{-1}L_1\rangle)^{-1}\langle P_0\rangle\| \ll 1$ and

(2) $(L_0 - \langle P\rangle - \langle L_1[L_0 - \langle P\rangle]^{-1}L_1\rangle - \langle P_0\rangle)^{-1}I\delta(x-x')\delta(t-t')$ is the leading term in the series (2.12),

then

$$\langle G(x,t|x',t')\rangle \approx \langle G_p(x,t|x',t')\rangle$$

and

$$\langle G_p(x,t|x',t')\rangle = (L_0 - \langle L_1[L_0 - \langle P\rangle]^{-1}L_1\rangle - \langle P\rangle - \langle P_0\rangle)^{-1}I\delta(x-x')\delta(t-t').$$

As we mentioned, the choice of the operator $\langle P_0\rangle$ depends on the physical situation being investigated, and its plausibility analysis is also predicated on the same condition.

A possible choice of $\langle P_0\rangle$ might be: (1) the *nearest neighbor* operator

$$\langle P_0\rangle \equiv \langle L_1[L_0 - \langle P\rangle]^{-1}L_1[L_0 - \langle P\rangle]^{-1}L_1[L_0 - \langle P\rangle]^{-1} \times L_1[L_0 - \langle P\rangle]^{-1}L_1\rangle \quad (2.14)$$

with the Green's equation

$$[L_0 - \langle P\rangle - \langle L_1[L_0 - \langle P\rangle]^{-1}L_1\rangle - \langle L_1[L_0 - \langle P\rangle]^{-1}L_1[L_0 - \langle P\rangle]^{-1} \times L_1[L_0 - \langle P\rangle]^{-1}L_1\rangle]\langle G_p(x,t|x',t')\rangle = I\delta(x-x')\delta(t-t') \quad (2.15)$$

or (2) the Kraichnan projection

$$\langle P_0\rangle \equiv \langle L_1(x,t)[L_0(x,t) - \langle P(x,t)\rangle]^{-1}L_1(x,t)\rangle + \langle L_1(x,t) \times \langle [L_0(x,t) - \langle P(x,t)\rangle + L_1(x,t)]^{-1}\rangle L_1(x,t)\rangle \quad (2.16)$$

with the corresponding equation being

$$[L_0(x,t) - \langle P(x,t)\rangle - \langle L_1(x,t)\langle [L_0(x,t) - \langle P(x,t)\rangle + L_1(x,t)]^{-1}\rangle L_1(x,t)\rangle\langle G_p(x,t|x',t')\rangle = I\delta(x-x')\delta(t-t'). \quad (2.17)$$

The solution for Eq. (2.14) is quite obvious. However, in Eq. (2.17) the result is achieved by producing a continuous fraction.

3. THE PROPERTIES OF THE OPERATOR $\langle P(x,t)\rangle$

At this point we shall consider what is the operator $\langle P(x,t)\rangle$, i.e., what properties can be attributed to it in order to ensure a proper choice of it. As we said before, if $\langle P(x,t)\rangle = 0$, then $f[\psi(x,t,\omega)] = f[\psi^H(x,t,\omega)]$ and that the projection physically is a measure of deviation of $f[\psi(x,t,\omega)]$ from $f[\psi^H(x,t,\omega)]$.

Let us now focus our attention on the equation

$$f[\psi(x,t,\omega)] = f[\psi^H(x,t,\omega)] + \langle P(x,t)\rangle L^{-1}(x,t,\omega)f[\psi(x,t,\omega)]. \quad (3.1)$$

Solving for $f[\psi(x,t,\omega)]$, one obtains

$$f[\psi(x,t,\omega)] = \psi_p(x,t,\omega) + [L(x,t,\omega) - \langle P(x,t)\rangle]^{-1} \times L(x,t,\omega)f[\psi^H(x,t,\omega)], \quad (3.2)$$

where $[L(x,t,\omega) - \langle P(x,t)\rangle]\psi_p(x,t,\omega) = 0$. The term $\psi_p(x,t,\omega)$ is related to the result for the stochastic

Green's function in the previous section. Since our initial condition is assumed to be independent of statistical variations, then

$$\psi_p(x,t,\omega)=\int\int dx'\,dt' G_p(x,t\,|\,x',t',\omega)\psi_{\rm int}(x'0). \tag{3.3}$$

If $\langle P\rangle \to 0$, we see

$$L(x,t,\omega)f[\psi(x,t,\omega)]=L(x,t,\omega)f[\psi^H(x,t,\omega)],$$

which implies $f[\psi(x,t,\omega)]\to f[\psi^H(x,t,\omega)]+\psi_0(x,t,\omega)$. Thus we see that $[L(x,t,\omega)-\langle P(x,t)\rangle]^{-1}L(x,t,\omega)$ is a measure of deviation of $f[\psi(x,t,\omega)]$ from $\psi_0(x,t,\omega)$ and if $[L(x,t,\omega)-\langle P(x,t)\rangle]^{-1}L(x,t,\omega)\ll 1$, then $f[\psi(x,t,\omega)]\to\psi_0(x,t,\omega)$, this further implies that $\langle P(x,t)\rangle \gg -L(x,t,\omega)$.

Returning now to (3.1) and iterating, we arrive at

$$f[\psi(x,t,\omega)]$$
$$=\{1+\langle P(x,t)\rangle L^{-1}(x,t,\omega)+\cdots\}f[\psi^H(x,t,\omega)] \tag{3.4a}$$
$$=\{1-\langle P(x,t)\rangle L^{-1}(x,t,\omega)\}^{-1}f[\psi^H(x,t,\omega)] \tag{3.4b}$$
$$=L(x,t,\omega)[L(x,t,\omega)-\langle P(x,t)\rangle]^{-1}f[\psi^H(x,t,\omega)]. \tag{3.4c}$$

From Eqs. (3.4c) and (3.2) we have

$$\{[L(x,t,\omega)-\langle P(x,t)\rangle]^{-1}L(x,t,\omega)-L(x,t,\omega)\times[L(x,t,\omega)-\langle P(x,t)\rangle]^{-1}\}f[\psi^H(x,t,\omega)]=\psi_p(x,t,\omega) \tag{3.5}$$

or the commutator

$$[[L(x,t,\omega)-\langle P(x,t)\rangle]^{-1},L(x,t,\omega)]\,f[\psi^H(x,t,\omega)]=\psi_p(x,t,\omega).$$

The commutator relationship can be transformed in terms of the Green's function and averaged to be

$$\langle[G_p(x,t\,|\,x',t')L(x',t')-L(x,t)G_p(x,t\,|\,x',t')]f[\psi^H]\rangle=\langle\psi_p(x',t')\rangle\delta(x-x')\delta(t-t'). \tag{3.6}$$

Thus when one solves the linear equation and defines the operator $\langle P_0(x,t)\rangle$, the commutator condition must be maintained to ensure the validity and proper treatment of the nonlinearity $f[\psi(x,t,\omega)]$.

At this point we are left to the averaging process, which is the third and final consideration.

4. THE EVALUATION OF THE CROSS CORRELATIONS

The averaging process can be separated only if the variations in two functions vary greatly from one another. This can only be determined by the parameter and functional relations of the physical problem.

However, once the proper averaging has been performed, this will automatically dictate certain criteria for the solution of the linear stochastic Green's function and properties of the operator $\langle P(x,t)\rangle$.

Let us now consider the case when the variations in the nonlinear functional are faster than those of the transport and linear processes, i.e., the correlation times in the nonlinear part are larger than the other processes. Thus Eq. (1.10) becomes

$$\langle\psi(x,t)\rangle=\int\int dx'\,dt'\langle G_p(x,t\,|\,x',t')\rangle\{\psi_{\rm int}(x',0)+g(x',t')\}-\int\int dx'\,dt'\langle P(x',t')\rangle\int\int dx''\,dt''\langle G_p(x,t\,|\,x',t')\times G_0(x',t'\,|\,x'',t'')\rangle\{\psi_{\rm int}(x'',0)+g(x'',t'')\}+\int\int dx'\,dt'\langle G_p(x,t\,|\,x',t')\rangle\langle f[\psi^H(x',t')]\rangle. \tag{4.1}$$

Thus we see that the evaluation of $\langle\psi(x,t)\rangle$ depends on the solution of the linear stochastic Green's function $\langle G_p(x,t\,|\,x',t')\rangle$, the averaging of $f[\psi^H(x,t)]$, and the cross correlation of the Green's function for the equation

$$L(x,t,\omega)G_0(x,t\,|\,x',t',\omega)=I\delta(x-x')\delta(t-t')$$

and the stochastic Green's function. Consequently, one must solve the unperturbed and perturbed Green's function, by using the renormalized projection operator technique, and then perform the average of the two Green's functions.

Thus, since

$$G_0(x,t\,|\,x',t',\omega)=\{1+[L_0(x,t)-\langle P^{(0)}(x,t)\rangle-\langle L_1(x,t)L_0^{-1}(x,t)\times L_1(x,t)\rangle]^{-1}\langle P^{(0)}(x,t)\rangle\}\langle\mathbf{G}_0(x,t\,|\,x',t')\rangle+\delta G(x,t\,|\,x',t',\omega) \tag{4.2a}$$

[where $\langle P^0(x,t)\rangle$ is the projection operator associated with the unperturbed operator[1]] and recalling that

$$G_p(x,t\,|\,x',t',\omega)=[1-[L_0(x,t)-\langle P(x,t)\rangle-\langle L_1(x,t)[L_0(x,t)-\langle P(x,t)\rangle]^{-1}L_1(x,t)\rangle]^{-1}\langle P_0(x,t)\rangle\times\langle\underline{G}(x',t'\,|\,x'',t'')\rangle+\delta G_p(x',t'\,|\,x'',t'',\omega). \tag{4.2b}$$

we see that

$$\langle G_p(x,t\,|\,x',t')G_0(x',t'\,|\,x'',t'')\rangle=[1-[L_0(x,t)-\langle P(x,t)\rangle-\langle L_1(x,t)\times[L_0(x,t)-\langle P(x,t)\rangle]^{-1}L_1(x,t)\rangle]^{-1}\langle P_0(x,t)\rangle]\times\langle\underline{G}(x,t\,|\,x',t')\rangle[1-L_0(x',t')-\langle P^{(0)}(x',t')\rangle-\langle L_1(x',t')L_0^{-1}(x',t')L_1(x',t')\rangle]^{-1}\langle P^{(0)}(x',t')\rangle\times\langle\underline{G}_0(x',t'\,|\,x'',t'')\rangle+\langle\delta G_p(x,t\,|\,x',t')\times\delta G_0(x',t'\,|\,x'',t'')\rangle. \tag{4.3}$$

The first term in (4.3) is the product of the solution of the mean Green's function which is the averaged product of the fluctuations. By the RNPO technique we generate a series of consecutive equations for $\langle G_p\rangle$ and δG. If we retain only the leading terms in the equations, i.e., we have the two conditions mentioned in the second section.

Then

$$\langle G_p(x,t\,|\,x',t')G_0(x',t'\,|\,x'',t'')\rangle=[L_0(x,t)-\langle P(x,t)\rangle-\langle L_1(x,t)[L_0(x,t)-\langle P(x,t)\rangle]^{-1}\times L_1(x,t)\rangle-\langle P_0(x,t)\rangle]^{-1}I\delta(x-x')\delta(t-t')[L_0(x',t')-\langle L_1(x',t')L_0^{-1}(x',t')L_1(x',t')\rangle-\langle P^0(x',t')\rangle]^{-1}\times I\delta(x'-x'')\delta(t-t')+\langle\delta G_p(x,t\,|\,x',t')\times\delta G_0(x',t'\,|\,x'',t'')\rangle \tag{4.4}$$

since

$$\delta G_p(x,t|x',t',\omega) = [1-\Omega_p(x,t)]^{-1}\Omega_p(x,t)\langle G_p(x,t|x',t')\rangle + \cdots \quad (4.5a)$$

and

$$\delta G_0(x',t'|x'',t'',\omega) = [1-\Omega_0(x,t)]^{-1}\Omega_0(x,t)\langle G_0(x',t'|x'',t'')\rangle + \cdots . \quad (4.5b)$$

Recalling the definition of $\Omega(x,t)$ from the second section and combining (4.5a) and (4.5b) followed by averaging, we arrive at

$$\begin{aligned}\langle \delta G_p(x,t|x',t')\delta G_0(x',t'|x'',t'')\rangle &\approx \sum_{N=0}^{\infty}\sum_{N'=0}^{\infty}\{\langle L_1(x,t)\Omega_p^N(x,t)L_1(x',t')\Omega_p(x',t)\rangle \\ &\quad - \langle L_1(x,t)\Omega_p(x,t)\rangle\langle L_1(x',t')\Omega_p^{N'}(x',t')\rangle\} \\ &\quad + [L_0(x,t)-\langle P(x,t)\rangle]^{-1}\langle G_p(x,t|x',t')\rangle \\ &\quad \times L_0^{-1}(x',t')\langle G_0(x',t'|x'',t'')\rangle . \end{aligned} \quad (4.6)$$

Retaining the first and leading term in the series, we conclude

$$\begin{aligned}\langle \delta G_p(x,t|x',t')\delta G_0(x',t'|x'',t'')\rangle &\approx [L_0(x,t)-\langle P(x,t)\rangle]^{-1}\int\int dx'\,dt'\langle G_p(x,t|x',t')\rangle \\ &\quad \times L_0^{-1}(x',t')\int\int dx''\,dt''\langle G_0(x',t'|x'',t'')\rangle \\ &\quad \times\langle L_1(x,t)L_1(x',t')\rangle . \end{aligned} \quad (4.7)$$

The solution of Eq. (1.1) has three types of terms: (1) a term due to the linear stochastic Green's function (a Dyson type equation), (2) a term due to the two-point stochastic Green's function (a Bethe–Saltpeter equation), and finally (3) a term due to the propagation on the nonlinear function $f[\psi^H(x,t,\omega)]$, where the argument is replaced by the solution of the linear equation and the deviation of the nonlinearity is directly proportional to the product of the projection operator and the two-point correlation function.

If we have the condition that

$$\begin{aligned}&\langle G_p(x,t|x',t')f[\psi^H(x',t')]\rangle \\ &\quad = \langle G_p(x,t|x',t')\rangle\langle f[\psi^H(x',t')]\rangle ,\end{aligned}$$

this further implies that we have the following commutator relation:

$$\begin{aligned}&\langle [\langle G_p(x,t|x',t')\rangle L(x',t') \\ &\qquad - L(x,t)\langle G_p(x,t|x',t')\rangle]f[\psi^H(x',t')]\rangle \\ &\quad = \langle \psi_p(x',t')\rangle\delta(x-x')\delta(t-t').\end{aligned}$$

CONCLUSION

Our results show that the choice of the operator $\langle P(x,t)\rangle$ is extremely crucial and that a series of conditions must be simultaneously met namely; for the linear stochastic Green's function we have

(1) $\|[L_0-\langle P\rangle-\langle L_1[L_0-\langle P\rangle]^{-1}L_1\rangle]^{-1}\langle P_0\rangle\| \ll 1$

and

(2) $[L_0-\langle P\rangle-\langle L_1[L_0-\langle P\rangle]^{-1}L_1\rangle-\langle P_0\rangle]^{-1}I\delta(x-x')\delta(t-t')$

is the leading term in the series (2.12). In the case of the nonlinear representation we have the commutator relation

$$\begin{aligned}&\langle [G_p(\overset{\circ}{x},\overset{\circ}{t}|x',t'),L(x,t)]f[\psi^H(x',t')]\rangle \\ &\quad = \langle \psi_p(x,t)\rangle\delta(x-x')\delta(t-t'),\end{aligned}$$

where the superscript $^\circ$ refers to the change of $x,t \to x',t'$ depending on the side L is with respect of G_p; and finally the condition concerning the two-point Green's function is that $\langle P\rangle$ is such that

$$\begin{aligned}&[L_0(x,t)-\langle P(x,t)\rangle]^{-1}\langle G_p(x,t|x',t')\rangle \\ &\quad \times L_0^{-1}(x,t)\langle G_0(x',t'|x'',t'')\rangle\langle L_1(x,t)L\langle(x',t')\rangle\end{aligned}$$

is the leading term in the series (4.6). Therefore, the choice of the operator $\langle P(x,t)\rangle$ must simultaneously incorporate the above conditions.

[1]V.A. LoDato, J. Math. Phys. 14, 340 (1973), Paper I.
[2]V.A. LoDato, J. Math. Phys. 15, 1740 (1974), Paper II.

Nonlinear canonical transformations and their representations in quantum mechanics

P. A. Mello* and M. Moshinsky*†

Instituto de Fisica, Universidad de México (UNAM), México 20, D.F.,Mexico
(Received 29 April 1975)

In the last few years an extensive literature has developed on linear canonical tranformations and their representation in quantum mechanics. Applications of these results have been made to clustering theory in nuclei, problems of accidental degeneracy, etc. In the present paper we wish to turn our attention to nonlinear canonical transformations. We show that by dealing with appropriate functions f_α ($\alpha = 1,...,2n$) of x_i, p_i ($i = 1,...,n$) rather than with these variables themselves, we can in principle set unambiguously the equations that determine the representation in quantum mechanics of the canonical transformation under study. This result holds when the old and new functions f_α have the same spectrum. We discuss specific examples when this last condition is satisfied: nonlinear canonical transformations in the radial variable that were obtained from projection of linear ones in higher-dimensional spaces; canonical transformations that take us from one Hamiltonian to another with the same spectrum, be this one continuous or discrete; canonical transformations that relate two sets of integrals of motion (which include the Hamiltonians) when we are dealing with phase spaces of dimensions higher than 2, etc. We discuss briefly, in the concluding section, the possibility of extending our analysis to canonical tranformations that do not conserve the spectrum of the relevant operators.

1. INTRODUCTION

It is an interesting historical fact that while canonical transformations were of considerable interest to physicists when they developed the foundations of quantum mechanics[1,2] in the 1920's, this interest wained after the objective was achieved. We are not referring to point transformations, such as those associated with the rotation or permutation of the coordinates, whose impact on the further development of quantum mechanics has been considerable,[3,4] but rather to *general* canonical transformations in which the new coordinates and momenta are functions of both the old.

The interest was rekindled occasionally in the next three decades when particular problems led to groups of canonical transformations. Among the examples we have the work of Fock[5] and Bargmann[6] on the $O(4)$ symmetry group of the Coulomb problem and the unitary group of symmetries of the harmonic oscillator originally noticed by Jauch and Hill.[7] A real renaissance on the subject does not start though until the extensive use by Elliott[8] of the $SU(3)$ symmetry of the three-dimensional oscillator in problems of nuclear structure. The latter symmetry is related with groups of linear canonical transformations[9] and these, together with their unitary representations in quantum mechanics, have been extensively discussed in the literature.[10–15] As indicated in Ref. 15 (where further information on the literature is also given) these developments have had a wide range of applications going from problems in nuclear clustering theory to the understanding of the groups responsible for accidental degeneracy.

While the situation with the representations of linear canonical transformations in quantum mechanics was clarified, it remained more obscure in the general case. For nonlinear canonical transformations several problems appear. First the differential operators, in the equations that determine the unitary representation in quantum mechanics,[12–15] may have fractional or negative powers and thus are undefined. Second, as x_i, p_i do not commute, the new coordinates and momenta $\bar{x}_i(\mathbf{x}, \mathbf{p})$, $\bar{p}_i(\mathbf{x}, \mathbf{p})$ that are functions of the old may not be well defined as quantum mechanical operators[16] if they contain products or other types of functions of *both* x_i, p_i. Third, and more important, the operators x_i, p_i and $\bar{x}_i$, $\bar{p}_i$ may have different spectra, as happens for example in the trivial point transformation $\bar{x} = x^2$. What is then the sense of a representation U which relates x_i, p_i with $\bar{x}_i$, $\bar{p}_i$ by the transformations[2]

$$\bar{x}_i = U x_i U^{-1}, \quad \bar{p}_i = U p_i U^{-1}, \tag{1.1}$$

implying that the spectra of the old and new variables are the same?

In the present work we do not intend to deal with the third problem though we will make some general remarks on it at the conclusion of this paper. We will discuss though the first and second problems. We begin by obtaining the equations that determine the representations of nonlinear canonical transformations. We then solve these equations in particular examples for which the third problem is not present, i.e., the new and old operators involved have the same spectrum. While these examples are very special, the approach to them is general enough to suggest what could be done in other cases.

2. THE EQUATIONS THAT DETERMINE THE UNITARY REPRESENTATION OF A CANONICAL TRANSFORMATION

Consider a classical system with n degrees of freedom, described by the n coordinates x_i and the n momenta p_i ($i = 1, \ldots, n$), which we shall put together into a single $2n$-component vector[12,13] z_α, $\alpha = 1, \ldots, 2n$,

$$\left.\begin{aligned} z_i &= x_i \\ z_{n+i} &= p_i \end{aligned}\right\} \; i = 1, \ldots, n. \tag{2.1}$$

The Poisson bracket between two functions $f(\mathbf{z})$, $g(\mathbf{z})$ can be written as

$$\{f,g\}=\frac{\partial f}{\partial z_\alpha}K_{\alpha\beta}\frac{\partial g}{\partial z_\beta}, \tag{2.2}$$

where we used the summation convention for repeated indices and

$$\|K_{\alpha\beta}\|=\begin{pmatrix}0 & I\\ -I & 0\end{pmatrix}, \tag{2.3}$$

with I being the n-dimensional unit matrix. In particular, the components z_α, z_β will satisfy the identify

$$\frac{\partial z_\alpha}{\partial z_\delta}K_{\delta\mu}\frac{\partial z_\beta}{\partial z_\mu}=K_{\alpha\beta}, \tag{2.4a}$$

or equivalently

$$\{x_i,p_j\}=\delta_{ij},\quad \{x_i,x_j\}=\{p_i,p_j\}=0. \tag{2.4b}$$

Now assume that we are given a set of $2n$ independent relations

$$f_\alpha(\mathbf{z})=\bar{f}_\alpha(\bar{\mathbf{z}}), \tag{2.5}$$

from which the $\bar{z}_\alpha$ can be obtained in terms of the f_α and hence of the z_α. We thus require that the Jacobians $|\partial f_\alpha/\partial z_\beta|$, $|\partial \bar{f}_\alpha/\partial \bar{z}_\beta|$ are not identically zero, so that (2.5) can be solved for the $\bar{z}_\alpha$ locally, even though the solution may not be unique. Suppose that we can make a definite choice among the various solutions, so that the new $\bar{z}_\alpha$ obey the Poisson bracket relations

$$\frac{\partial \bar{z}_\alpha}{\partial z_\beta}K_{\delta\mu}\frac{\partial \bar{z}_\beta}{\partial z_\mu}=K_{\alpha\beta} \tag{2.6}$$

with respect to the old z_α, i.e., that the transformation is canonical. We shall find a condition that the f_α's must fulfill, so that (2.6) are satisfied. We can write (2.6) as

$$\frac{\partial \bar{z}_\alpha}{\partial \bar{f}_{\gamma'}}\frac{\partial f_{\gamma'}}{\partial z_\delta}K_{\delta\mu}\frac{\partial \bar{z}_\beta}{\partial \bar{f}_{\lambda'}}\frac{\partial f_{\lambda'}}{\partial z_\mu}=K_{\alpha\beta}, \tag{2.7}$$

from which we obtain

$$\frac{\partial \bar{f}_\gamma}{\partial \bar{z}_\alpha}\frac{\partial \bar{z}_\alpha}{\partial \bar{f}_{\gamma'}}\frac{\partial f_{\gamma'}}{\partial z_\delta}K_{\delta\mu}\frac{\partial \bar{z}_\beta}{\partial \bar{f}_{\lambda'}}\frac{\partial f_{\lambda'}}{\partial z_\mu}\frac{\partial \bar{f}_\lambda}{\partial \bar{z}_\beta}=\frac{\partial \bar{f}_\gamma}{\partial \bar{z}_\alpha}K_{\alpha\beta}\frac{\partial \bar{f}_\lambda}{\partial \bar{z}_\beta}. \tag{2.8}$$

Since

$$\frac{\partial \bar{f}_\gamma}{\partial \bar{z}_\alpha}\frac{\partial \bar{z}_\alpha}{\partial \bar{f}_{\gamma'}}=\delta_{\gamma\gamma'},\quad \frac{\partial \bar{z}_\beta}{\partial \bar{f}_{\lambda'}}\frac{\partial \bar{f}_\lambda}{\partial \bar{z}_\beta}=\delta_{\lambda\lambda'}, \tag{2.9}$$

we have

$$\frac{\partial f_\gamma}{\partial z_\delta}K_{\delta\mu}\frac{\partial f_\lambda}{\partial z_\mu}=\frac{\partial \bar{f}_\gamma}{\partial \bar{z}_\alpha}K_{\alpha\beta}\frac{\partial \bar{f}_\lambda}{\partial \bar{z}_\beta}, \tag{2.10}$$

or

$$\{f_\gamma,f_\lambda\}_z=\{\bar{f}_\gamma,\bar{f}_\lambda\}_{\bar{z}}. \tag{2.11}$$

Therefore, a necessary condition for the transformation (2.5) to be canonical, is that the Poisson bracket between f_γ and f_λ with respect to the z_α be the same as that between $\bar{f}_\gamma$ and $\bar{f}_\lambda$ with respect to the $\bar{z}_\alpha$, when use is made of (2.5) to substitute the $\bar{z}_\alpha$ in terms of the z_α. That this condition is also sufficient is clear, since the steps leading from (2.6) to (2.11) can all be taken in the reversed order.

We shall now discuss the corresponding quantum mechanical problem, obtained when the classical quantities x_i, p_i, $\bar{x}_i$, $\bar{p}_i$ in (2.5) are regarded as operators. In dealing with the quantum problem, we shall follow Dirac's convention[2] of attaching primes to those quantities that are interpreted as c-numbers. We shall also distinguish between the eigenstates of the old and new operators by means of angular and round brackets, respectively,[12–15] e.g.,

$$x_i|\mathbf{x}'\rangle=x_i'|\mathbf{x}'\rangle,\quad \bar{x}_i|\bar{\mathbf{x}}'\rangle=\bar{x}_i'|\bar{\mathbf{x}}'\rangle, \tag{2.12a}$$

$$p_i|\mathbf{p}'\rangle=p_i'|\mathbf{p}'\rangle,\quad \bar{p}_i|\bar{\mathbf{p}}'\rangle=\bar{p}_i'|\bar{\mathbf{p}}'\rangle. \tag{2.12b}$$

The problem of finding, in quantum mechanics, the unitary representation of the classical canonical transformation (2.5) is thus that of finding the operator U that relates the barred and unbarred operators, or the round and angular brackets:

$$\bar{x}_i=Ux_iU^{-1},\quad \bar{p}_i=Up_iU^{-1}, \tag{2.13a}$$

$$|\mathbf{x}'\rangle=U|\mathbf{x}'\rangle,\quad |\mathbf{p}'\rangle=U|\mathbf{p}'\rangle. \tag{2.13b}$$

In the coordinate representation, the matrix elements of U can thus be written as

$$\langle x'|U|\mathbf{x}'\rangle=\langle x'|\mathbf{x}'\rangle. \tag{2.14}$$

In terms of (2.5), considered now as operator relations, U must satisfy relations analogous to (2.13a), i.e.,

$$\bar{f}_\alpha(\bar{\mathbf{x}},\bar{\mathbf{p}})=U\bar{f}_\alpha(\mathbf{x},\mathbf{p})\,U^{-1}=f_\alpha(\mathbf{x},\mathbf{p}), \tag{2.15}$$

or

$$\bar{f}_\alpha(\bar{\mathbf{x}},\bar{\mathbf{p}})\,U=U\bar{f}_\alpha(\mathbf{x},\mathbf{p}). \tag{2.16}$$

As was mentioned in the Introduction, relations (2.13a) and (2.15) only make sense if the operators $\bar{x}_i$ and x_i have the same spectrum, with this same condition on the pairs $\bar{p}_i$, p_i and $\bar{f}_\alpha$, f_α. If these requirements were not fulfilled, i.e., if the third problem mentioned in the Introduction appeared, perhaps the appropriate starting point could be Eq. (2.16) instead of (2.15), with no unitarity condition on the operator U; in fact, U might not even have an inverse.

Going back to (2.16), we introduce the unit operator $\int|\mathbf{x}''\rangle\,d\mathbf{x}''\langle\mathbf{x}''|$ between $\bar{f}_\alpha$ and U and take the matrix element of both sides of (2.16) between the states $\langle\mathbf{x}'|$ and $|\bar{\mathbf{x}}'\rangle$, which are both eigenstates of the *old* position operator.

We thus have

$$\int\langle\mathbf{x}'|\bar{f}_\alpha(\bar{\mathbf{x}},\bar{\mathbf{p}})|\mathbf{x}''\rangle\,d\mathbf{x}''\langle\mathbf{x}''|U|\bar{\mathbf{x}}'\rangle$$

$$=\int\langle\mathbf{x}'|U|\mathbf{x}''\rangle\,d\mathbf{x}''\langle\mathbf{x}''|\bar{f}_\alpha(\mathbf{x},\mathbf{p})|\bar{\mathbf{x}}'\rangle. \tag{2.17}$$

In the left-hand side we write the matrix elements of $\bar{f}_\alpha$ in terms of the old variables x_i', $-i\partial/\partial x_i'$, and using the last equality in (2.15)

$$\langle\mathbf{x}'|\bar{f}_\alpha(\bar{\mathbf{x}},\bar{\mathbf{p}})|\mathbf{x}''\rangle=f_\alpha\left(\mathbf{x}',-i\frac{\partial}{\partial\mathbf{x}'}\right)\delta(\mathbf{x}'-\mathbf{x}''). \tag{2.18}$$

Recalling also, from (2.14), that

$$\langle\mathbf{x}''|U|\bar{\mathbf{x}}'\rangle=\langle\mathbf{x}''|\bar{\mathbf{x}}'\rangle, \tag{2.19}$$

the left-hand side of (2.17) becomes

$$\int f_\alpha\left(\mathbf{x}',-i\frac{\partial}{\partial\mathbf{x}'}\right)\delta(\mathbf{x}'-\mathbf{x}'')\,d\mathbf{x}''\langle\mathbf{x}''|\bar{\mathbf{x}}'\rangle$$

$$=f_\alpha\left(\mathbf{x}',-i\frac{\partial}{\partial\mathbf{x}'}\right)\langle\mathbf{x}'|\bar{\mathbf{x}}'\rangle. \tag{2.20}$$

Concerning the right-hand side of (2.17), we first write it as

$$\int \langle \mathbf{x}' | U | \mathbf{x}'' \rangle d\mathbf{x}'' \langle \mathbf{x}'' | \bar{f}_\alpha(\mathbf{x}, \mathbf{p}) | \bar{\mathbf{x}}' \rangle = [\int \langle \bar{\mathbf{x}}' | \bar{f}_\alpha^\dagger(\mathbf{x}, \mathbf{p}) | \mathbf{x}'' \rangle d\mathbf{x}'' \langle \mathbf{x}'' | U^\dagger | \mathbf{x}' \rangle]^*, \quad (2.21)$$

where we used the relations:

$$U_{ij} F_{jk} = (UF)_{ik} = [(UF)^\dagger_{ki}]^* = [(F^\dagger U^\dagger)_{ki}]^* = (F^\dagger_{kj} U^\dagger_{ji})^*. \quad (2.22)$$

We can then, using the notation (2.14), write (2.21) as:

$$\left[\int \bar{f}_\alpha^\dagger\left(\bar{\mathbf{x}}', -i\frac{\partial}{\partial \bar{\mathbf{x}}'}\right) \delta(\bar{\mathbf{x}}' - \mathbf{x}'') d\mathbf{x}'' \langle \mathbf{x}'' | \mathbf{x}' \rangle \right]^* = \left[\bar{f}_\alpha^\dagger\left(\bar{\mathbf{x}}', -i\frac{\partial}{\partial \bar{\mathbf{x}}'}\right) \langle \bar{\mathbf{x}}' | \mathbf{x}' \rangle \right]^* = \left[\bar{f}_\alpha^\dagger\left(\bar{\mathbf{x}}', -i\frac{\partial}{\partial \mathbf{x}'}\right) \right]^* \langle \mathbf{x}' | \bar{\mathbf{x}}' \rangle. \quad (2.23)$$

Equating (2.20) and (2.23), we obtain

$$f_\alpha\left(\mathbf{x}', -i\frac{\partial}{\partial \mathbf{x}'}\right) \langle \mathbf{x}' | \bar{\mathbf{x}}' \rangle = \left[\bar{f}_\alpha^\dagger\left(\bar{\mathbf{x}}', -i\frac{\partial}{\partial \bar{\mathbf{x}}'}\right) \right]^* \langle \mathbf{x}' | \bar{\mathbf{x}}' \rangle, \quad \alpha = 1, \ldots, 2n, \quad (2.24)$$

which form a set of $2n$ coupled differential equations for the transformation bracket $\langle \mathbf{x}' | \bar{\mathbf{x}}' \rangle$.

Sometimes the relations between $\bar{x}_i$, $\bar{p}_i$ and x_i, p_i is sufficiently simple that one can work with them directly, giving $2n$ differential equations of the type[12,13]

$$\bar{x}_i\left(\mathbf{x}', -i\frac{\partial}{\partial \mathbf{x}'}\right) \langle \mathbf{x}' | \bar{\mathbf{x}}' \rangle = \bar{x}_i' \langle \mathbf{x}' | \mathbf{x}' \rangle, \quad i = 1, \ldots, n. \quad (2.25a)$$

$$\bar{p}_i\left(\mathbf{x}', -i\frac{\partial}{\partial \mathbf{x}'}\right) \langle \mathbf{x}' | \bar{\mathbf{x}}' \rangle = i\frac{\partial}{\partial \mathbf{x}_i'} \langle \mathbf{x}' | \bar{\mathbf{x}}' \rangle. \quad (2.25b)$$

On occasions, however, it might prove advantageous to use combinations f_α of the variables x_i, p_i, which might be more convenient to handle, especially as far as the corresponding quantization is concerned. If we can find combinations f_α for which the first and second problems mentioned in the introduction do not appear, then it is desirable to use the general equations (2.24) instead of (2.25).

Particular applications of the ideas which we have just discussed will be considered in the following sections.

3. NONLINEAR CANONICAL TRANSFORMATIONS IN RADIAL PHASE SPACE

The group of canonical transformations we wish to consider comes from the linear transformation

$$\bar{r} = ar + b\mathbf{p}, \quad \bar{\mathbf{p}} = cr + d\mathbf{p}, \quad ad - bc = 1, \quad (3.1)$$

where all the vectors are two dimensional. Taking into account that the angular momentum[17] remains invariant under the transformation (3.1) and denoting it by λ, we have for the radial coordinate and momenta (which instead of r, p_r we again call x, p) the classical canonical transformation[17]

$$\bar{x} = [a^2x^2 + b^2(p^2 + \lambda^2 x^{-2}) + 2ab\,xp]^{1/2}, \quad (3.2a)$$

$$\bar{p} = \frac{ac\,x^2 + bd(p^2 + \lambda^2 x^{-2}) + (ad + bc)xp}{[a^2x^2 + b^2(p^2 + \lambda^2 x^{-2}) + 2ab\,xp]^{1/2}}. \quad (3.2b)$$

To determine the representation $\langle \mathbf{x}' | \bar{\mathbf{x}}' \rangle$ for the case (3.2) we proceeded, in Ref. 17, by the trick of projecting from the unitary representation of the linear canonical transformation (3.1) in four-dimensional phase space. If we had wanted to use Eq. (2.25) directly we immediately face two problems. One, which is simple, is how to interpret products such as xp as operators, which is resolved if we use the Hermitian from $(xp + px)/2$. The other concerns the fact that we have square roots and inverses of differential operators. For the latter problem we note that instead of the operators in (2.25) we could have used some function of them. For example consider Eq. (2.25a) and apply to both sides of it the operator $\bar{p}$ taking into account that, as a function $x', -i\partial/\partial x'$ only, it commutes with $\bar{x}'$. Then using (2.25b) we have

$$\bar{p}\left(x', -i\frac{\partial}{\partial x'}\right) \bar{x}\left(x', -i\frac{\partial}{\partial x'}\right) \langle x' | \bar{x}' \rangle = i\bar{x}' \frac{\partial}{\partial \bar{x}'} \langle x' | \bar{x}' \rangle. \quad (3.3)$$

A similar result holds when we apply the operator $\bar{x}$ to both sides of (2.25a). Symmetrizing Eq. (3.3) we see that the unitary representation $\langle x' | \bar{x}' \rangle$ satisfies from (2.25) and (3.2) the equations

$$\left[a^2x'^2 + b^2\left(-\frac{\partial^2}{\partial x'^2} + \frac{\lambda^2}{x'^2}\right) - iab\left(x'\frac{\partial}{\partial x'} + \frac{\partial}{\partial x'}x'\right)\right] \langle x' | \bar{x}' \rangle = \bar{x}'^2 \langle x' | \bar{x}' \rangle, \quad (3.4a)$$

$$\left[acx'^2 + bd\left(-\frac{\partial^2}{\partial x'^2} + \frac{\lambda^2}{x'^2}\right) - \frac{i}{2}(ad + bc)\left(x'\frac{\partial}{\partial x'} + \frac{\partial}{\partial x'}x'\right)\right] \times \langle x' | \bar{x}' \rangle = \frac{i}{2}\left(\bar{x}'\frac{\partial}{\partial \bar{x}'} + \frac{\partial}{\partial \bar{x}'}\bar{x}'\right) \langle x' | \bar{x}' \rangle, \quad (3.4b)$$

$$\int_0^\infty \langle \bar{x}' | x' \rangle dx' \langle x' | \bar{x}'' \rangle = \delta(\bar{x}' - \bar{x}''), \quad (3.4c)$$

where (3.4c) is due to the fact that the spectrum of both $x', \bar{x}'$ goes from 0 to ∞. We could also have arrived at Eq. (3.4) if we had started with the hermitized form of the operators $\bar{x}^2$, $\bar{x}\bar{p}$ and employed the general equations (2.24).

Equations (3.4a, b) are then a set of two well-defined partial differential equations in $x', \bar{x}'$. To solve them we multiply the first by d and the second by b and subtract. Using the relation $ad - bc = 1$, we arrive at the equations

$$\left(ax'^2 - ibx'\frac{\partial}{\partial x'}\right) \langle x' | \bar{x}' \rangle = \left(d\bar{x}'^2 - ib\bar{x}'\frac{\partial}{\partial \bar{x}'}\right) \langle x' | \bar{x}' \rangle. \quad (3.5)$$

Proposing a solution of (3.5) of the form

$$\langle x' | \bar{x}' \rangle = f(x', \bar{x}') \exp[-i(2b)^{-1}(ax'^2 + d\bar{x}'^2)], \quad (3.6)$$

we obtain that

$$x'\frac{\partial}{\partial x'} f(x', \bar{x}') = \bar{x}'\frac{\partial}{\partial \bar{x}'} f(x', \bar{x}'), \quad (3.7)$$

and thus f is only a function of the product of these

variables, i.e., $f(x'\bar{x}')$. Substituting in (3.4a) we get finally that $f(x'\bar{x}')$ satisfies the equation

$$\left(-b^2\frac{\partial^2}{\partial x'^2}+b^2\frac{\lambda^2}{x'^2}-\bar{x}'^2\right)f(x'\bar{x}')=0, \tag{3.8}$$

which has the immediate solution

$$f(x'\bar{x}')=A(x'\bar{x}')^{1/2}\,J_\mu(b^{-1}x'\bar{x}'), \tag{3.9}$$

where A is an arbitrary constant, J_μ is a Bessel function and

$$\mu=(\lambda^2+\tfrac{1}{4})^{1/2}. \tag{3.10}$$

We still have to satisfy the normalization condition (3.4c) which is easily achieved[18] by selecting A as b^{-1} and thus $\langle x'|\bar{x}'\rangle$ becomes

$$\langle x'|\bar{x}'\rangle=b^{-1}(x'\bar{x}')^{1/2}J_\mu(b^{-1}x'\bar{x}')\exp\left(-\frac{i}{2b}(ax'^2+d\bar{x}'^2)\right), \tag{3.11}$$

which coincides with the result of Ref. 17 obtained by projection from the larger dimensional representation of the linear canonical transformation (3.1).

The replacement of Eqs. (2.25) with the operators $\bar{x}$, $\bar{p}$ given by (3.2), by the Eqs. (2.24) associated with the hermitized form of x^2, $\bar{x}\bar{p}$ lead to the "bonafide" set of partial differential equations (3.4). For the latter, problems of the type one and two mentioned in the Introduction (fractional or negative powers of the differential operators or order of factors) disappear. In fact, we obtained the explicit solution (3.11) of Eqs. (3.4).

The procedure followed in this section indicates the importance of using functions of $\bar{x}$, $\bar{p}$, rather than the observables themselves, in obtaining the equations that determine the representation of a canonical transformation. In the next section we apply this approach to canonical transformations in two-dimensional phase space that take us from one Hamiltonian to another.

4. REPRESENTATIONS OF CANONICAL TRANSFORMATIONS RELATING DIFFERENT HAMILTONIANS WITH THE SAME CONTINUOUS SPECTRUM

Let us consider two classical Hamiltonians H and $\bar{H}$ associated respectively with the phase space variables x, p and $\bar{x}$, $\bar{p}$ by the equations

$$H=\tfrac{1}{2}p^2+V(x),\quad \bar{H}=\tfrac{1}{2}\bar{p}^2+\bar{V}(\bar{x}), \tag{4.1a,b}$$

where we take the mass as unity. What is a canonical transformation that takes H into $\bar{H}$? We start by considering the variables canonically conjugate to $H,\bar{H}$; they are clearly the time[19] and as classically $p=\dot{x}$, $\bar{p}=\dot{\bar{x}}$, we get

$$T=\int_a^x\frac{dy}{[2H-2V(y)]^{1/2}},\quad \bar{T}=\int_{\bar{a}}^{\bar{x}}\frac{dy}{[2H-2\bar{V}(y)]^{1/2}}, \tag{4.2a,b}$$

where y is an integration variable that allows us to determine T as a function of x and the H of (4.1a) and similarly for $\bar{T}$. The $a,\bar{a}$ are arbitrary functions of H. We easily check that

$$\{T,H\}_{x,p}=\{\bar{T},\bar{H}\}_{\bar{x},\bar{p}}=1. \tag{4.3}$$

If we now establish the equations

$$T(x,p)=\bar{T}(\bar{x},\bar{p}),\quad H(x,p)=\bar{H}(\bar{x},\bar{p}), \tag{4.4a,b}$$

we get $\bar{x}$, $\bar{p}$ and x, p related by a canonical transformation,[19] which furthermore is, by construction, the one that takes $H(x,p)$ into $\bar{H}(\bar{x},\bar{p})$.

The classical part being clear we now turn to the quantum picture, i.e., the determination of $\langle x'|\bar{x}'\rangle$. To avoid the problems of the third type that were mentioned in the Introduction we shall consider Hamiltonians $H,\bar{H}$ that have the same spectrum E which, in this section, we assume continuous and restricted by $0\le E<\infty$.

In Fig. 1 we show two potentials, indicated by the solid and dotted lines that give the type of spectrum we demand. Both $V(x)$ and $\bar{V}(x)$ are positive for all x, vanish at infinity and become infinite at $x=0$. The last restriction has as its purpose that eigenstates of $H,\bar{H}$ should be nondegenerate. If $V(x)$, $\bar{V}(x)$ are positive, regular at $x=0$, and vanish at $\pm\infty$, the states are doubly degenerate depending on whether the wavefunction propagates to the left or right. This last case requires only a trivial extension of the reasoning that we shall proceed to implement for potentials of the type shown in Fig. 1, extension that we will indicate at the end of this section.

Before proceeding to derive the equations that determine the unitary representation associated with the canonical transformation (4.4a,b) we note that we could replace (4.4a) by the equation

$$K(H,T)=K(\bar{H},\bar{T}), \tag{4.4c}$$

where K is some arbitrary function. Clearly in the classical case Eqs. (4.4a,b) and (4.4b,c) lead to the same canonical transformation $\bar{x}=\bar{x}(x,p)$, $\bar{p}=\bar{p}(x,p)$. But in quantum mechanics, as will be seen in the examples discussed in the next sections, $K(H,T)$ may be a function of x,p that does not present problems of the

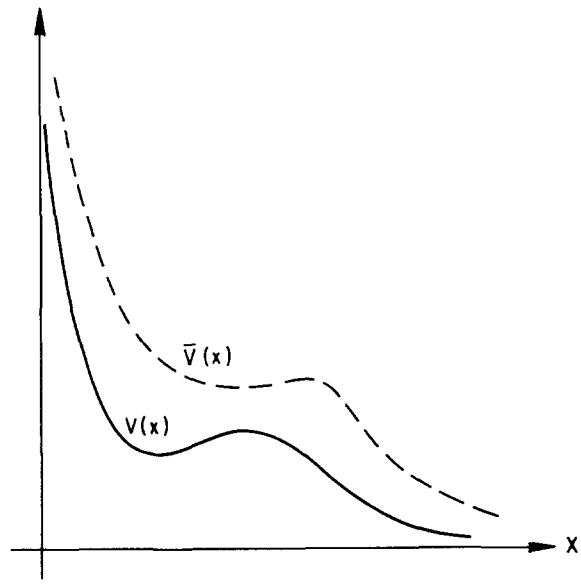


FIG. 1. The potentials $V(x)$ and $\bar{V}(x)$ that define the two Hamiltonians H, $\bar{H}$ to be related by a canonical transformation.

type one and two that were mentioned in the Introduction. Note that $K(H, T)$, $K(\bar{H}, \bar{T})$ are functions of x, p and $\bar{x}$, $\bar{p}$, respectively, which we could write as $G(x, p)$, $\bar{G}(\bar{x}, \bar{p})$. Equations (4.4c) then take the form

$$G(x, p) = \bar{G}(\bar{x}, \bar{p}). \tag{4.4d}$$

We can finally state that the classical canonical transformation is determined by the two equations (4.4b, d).

The unitary representation $\langle x' | \bar{x}' \rangle$ associated with the canonical transformation defined by (4.4b, d), for potentials of the form given in Fig. 1, satisfies then equations of the type (2.24) which now take the form

$$H\left(x', -i\frac{\partial}{\partial x'}\right) \langle x' | \bar{x}' \rangle = \bar{H}\left(\bar{x}, -i\frac{\partial}{\partial \bar{x}'}\right) \langle x' | \bar{x}' \rangle, \tag{4.5a}$$

$$G\left(x', -i\frac{\partial}{\partial x'}\right) \langle x' | \bar{x}' \rangle = \bar{G}^*\left(\bar{x}', -i\frac{\partial}{\partial \bar{x}'}\right) \langle x' | \bar{x}' \rangle, \tag{4.5b}$$

$$\int_0^\infty \langle \bar{x}' | x' \rangle dx' \langle x' | \bar{x}'' \rangle = \delta(\bar{x}' - \bar{x}''). \tag{4.5c}$$

In (4.5) the Hamiltonian H is both Hermitian and real and thus $H^{\dagger *} = H$, while we take the function G as Hermitian (or if it is not so originally we hermitize it) and therefore $G^{\dagger *} = G^*$.

How can we proceed to solve Eqs. (4.5) in a general way? First we look for the normalized eigenfunctions of the Hamiltonians

$$H\left(x', -i\frac{\partial}{\partial x'}\right) \psi_E(x') = E\psi_E(x'), \tag{4.6a}$$

$$\bar{H}\left(\bar{x}', -i\frac{\partial}{\partial \bar{x}'}\right) \bar{\psi}_E(\bar{x}') = E\bar{\psi}_E(\bar{x}'), \tag{4.6b}$$

$$\int_0^\infty \psi_E^*(x')\, \psi_{E'}(x')\, dx' = \delta(E - E'), \tag{4.6c}$$

$$\int_0^\infty \bar{\psi}_E^*(\bar{x}')\, \bar{\psi}_{E'}(\bar{x}')\, d\bar{x}' = \delta(E - E'). \tag{4.6d}$$

Furthermore, from the completeness of these eigenfunctions we must have the following decomposition of the unit operator

$$\int_0^\infty \psi_E(x')\, \psi_E^*(x'')\, dE = \delta(x' - x''), \tag{4.6e}$$

$$\int_0^\infty \bar{\psi}_E(\bar{x}')\bar{\psi}_E^*(\bar{x}'')\, dE = \delta(\bar{x}' - \bar{x}''). \tag{4.6f}$$

From Eqs. (4.6) we see then immediately that if we write

$$\langle x' | \bar{x}' \rangle = \int_0^\infty \exp[i\phi(E)]\, \psi_E(x')\, \bar{\psi}_E^*(\bar{x}')\, dE, \tag{4.7}$$

where $\phi(E)$ is a real as yet undetermined function of E, Eqs. (4.5a, c) are satisfied. It remains then to satisfy (4.5b) which should determine the phase factor ϕ in (4.7). We can then write (4.5b) as

$$\int_0^\infty \exp[i\phi(E)] \left\{ \left[G\left(x', -i\frac{\partial}{\partial x'}\right) \psi_E(x') \right] \bar{\psi}_E^*(\bar{x}') - \psi_E(x') \left[G\left(\bar{x}', -i\frac{\partial}{\partial \bar{x}'}\right) \bar{\psi}_E(\bar{x}') \right]^* \right\} dE = 0. \tag{4.8}$$

Equation (4.8) is a complicated transform of $\exp[i\phi(E)]$ that must vanish for all $x', \bar{x}'$. In principle it should determine the $\phi(E)$ though not necessarily in a unique fashion. Once this is achieved the integral (4.7) can be evaluated leading to $\langle x' | \bar{x}' \rangle$. We shall implement this program explicitly in the next section for centrifugal potentials of different strengths.

As a last point we indicate that if $V(x)$, $\bar{V}(x)$ are positive, but regular at the origin and vanish at $\pm\infty$, we must consider eigenfunctions $\psi_E^{(\pm)}(x')$, $\psi_E^{(\pm)}(\bar{x}')$ associated with the fact that the solutions represent waves going to the right or to the left. We have then also a corresponding phase $\phi_\pm(E)$, and $\langle x' | \bar{x}' \rangle$ is given by the sum of two terms of the type (4.7), one with a + and the other with a − sign.

5. CANONICAL TRANSFORMATIONS RELATING CENTRIFUGAL POTENTIALS OF DIFFERENT STRENGTHS

In this section we consider the problem where the Hamiltonians $H, \bar{H}$ are given by

$$H = \tfrac{1}{2}p^2 + \frac{\lambda^2}{2x^2}, \tag{5.1a}$$

$$\bar{H} = \tfrac{1}{2}\bar{p}^2 + \frac{\bar{\lambda}^2}{2\bar{x}^2}, \tag{5.1b}$$

where $\lambda, \bar{\lambda}$ are two arbitrary real numbers. According to (4.2) the corresponding $T, \bar{T}$ will be given by

$$T = (2H)^{-1}(2Hx^2 - \lambda^2)^{1/2} = (2H)^{-1} xp, \tag{5.2a}$$

$$\bar{T} = (2\bar{H})^{-1}\bar{x}\bar{p}, \tag{5.2b}$$

where we have taken $a, \bar{a}$ so that their contribution to (4.2) is zero. We easily check that $\{T, H\}_{x,p} = \{\bar{T}, \bar{H}\}_{\bar{x},\bar{p}} = 1$ and furthermore we can introduce the function

$$K(H, T) \equiv G(x, p) = 2HT = xp, \tag{5.3a}$$

$$K(\bar{H}, \bar{T}) \equiv G(\bar{x}, \bar{p}) = 2\bar{H}\bar{T} = \bar{x}\bar{p}. \tag{5.3b}$$

The classical canonical transformation that takes us from H to $\bar{H}$ is, from (4.4b, d), given then by the equations

$$\tfrac{1}{2}p^2 + \frac{\lambda^2}{2x^2} = \tfrac{1}{2}\bar{p}^2 + \frac{\bar{\lambda}^2}{2\bar{x}^2}, \tag{5.4a}$$

$$xp = \bar{x}\bar{p}, \tag{5.4b}$$

which lead to

$$\bar{x} = x\left(\frac{x^2p^2 + \bar{\lambda}^2}{x^2p^2 + \lambda^2}\right)^{1/2}, \tag{5.5a}$$

$$\bar{p} = p\left(\frac{x^2p^2 + \lambda^2}{x^2p^2 + \bar{\lambda}^2}\right)^{1/2}, \tag{5.5b}$$

where again we check that $\{\bar{x}, \bar{p}\}_{x,p} = 1$.

We wish now to find the unitary representation $\langle x' | \bar{x}' \rangle$ corresponding to the nonlinear canonical transformation (5.5a, b). We shall achieve this in two ways, one by using an expansion of the type (4.7) and the other by directly solving Eqs. (4.5).

For the first type of approach we need the eigenstates of the Hamiltonians (5.1). Writing

$$E = \tfrac{1}{2}k^2, \quad \mu = (\lambda^2 + \tfrac{1}{4})^{1/2}, \quad \bar{\mu} = (\bar{\lambda}^2 + \tfrac{1}{4})^{1/2}, \tag{5.6}$$

these become, in the notation of (4.6),

$$\psi_E(x') = (x')^{1/2} J_\mu(kx'), \tag{5.7a}$$

$$\bar{\psi}_E(\bar{x}') = (\bar{x}')^{1/2} J_{\bar{\mu}}(k\bar{x}'), \tag{5.7b}$$

where the J_μ's are Bessel functions of the order indicated. The integral (4.6c) is then[18]

$$\int_0^\infty \psi_E^*(x')\,\psi_{E'}(x')\,dx'$$
$$= \int_0^\infty x' J_\mu(kx') J_\mu(k'x')\,dx' = (kk')^{-1/2}\,\delta(k-k')$$
$$= \delta(E-E'), \tag{5.8}$$

and similarly for $\bar{\psi}_{\bar{E}}(\bar{x}')$, so that the functions are normalized. We can thus, from (4.7), write

$$\langle x'|\bar{x}'\rangle = \int_0^\infty \exp[i\chi(k)]\,(kx')^{1/2} J_\mu(kx')(k\bar{x}')^{1/2} J_{\bar{\mu}}(k\bar{x}')\,dk, \tag{5.9}$$

where we replaced E by $\frac{1}{2}k^2$ and $\phi(E)=\chi(k)$. It remains then to use (4.5b), which in the Hermitian operator form says

$$\frac{1}{2i}\left(x'\frac{\partial}{\partial x'} + \frac{\partial}{\partial x'}x'\right)\langle x'|\bar{x}'\rangle$$
$$= -\frac{1}{2i}\left(\bar{x}'\frac{\partial}{\partial \bar{x}'} + \bar{x}'\frac{\partial}{\partial \bar{x}'}\right)\langle x'|\bar{x}'\rangle \tag{5.10}$$

as $G(x,p)=xp$, $\bar{G}(\bar{x},\bar{p})=\bar{x}\bar{p}$.

Applying (5.10) to (5.9) and noticing that the effect of $x'\partial/\partial x'$ on $(kx')^{1/2}J_\mu(kx')$ is the same as $k\partial/\partial k$, we obtain

$$-i\left(x'\frac{\partial}{\partial x'} + \bar{x}'\frac{\partial}{\partial \bar{x}'} + 1\right)\int_0^\infty \exp[i\chi(k)]$$
$$\times (kx')^{1/2} J_\mu(kx')(k\bar{x}')^{1/2} J_{\bar{\mu}}(k\bar{x}')\,dk$$
$$= -i\int_0^\infty \exp[i\chi(k)]\left[\left(k\frac{\partial}{\partial k}+1\right)\right.$$
$$\left.\times (kx')^{1/2} J_\mu(kx')(k\bar{x}')^{1/2} J_{\bar{\mu}}(k\bar{x}')\right]dk$$
$$= -\int_0^\infty \frac{\partial\chi(k)}{\partial k}\exp[i\chi(k)]$$
$$\times (kx')^{1/2} J_\mu(kx')(k\bar{x}')^{1/2} J_{\bar{\mu}}(k\bar{x}')\,k\,dk = 0. \tag{5.11}$$

The last integral was derived by integration by parts, which is allowed in the sense of distribution theory, as the Bessel functions for $k\to\infty$ oscillate quickly and thus average out to zero in any integral.

Equation (5.11) will then be satisfied if $\partial\chi(k)/\partial k = 0$, implying that $\chi(k)$ is a constant which, without loss of generality (our representations are ray representations[14]), we can take as

$$\chi(k) = 0. \tag{5.12}$$

Introducing this value in (5.9) we find that the resulting integral, which is a Hankel transform,[20] is tabulated for $x'\neq\bar{x}'$. When $x'=\bar{x}'$ we must use the asymptotic form of the Bessel functions to determine the δ function behavior at that point. The full discussion is given in the Appendix and here we only present the final expression

$$\langle x'|\bar{x}'\rangle \equiv \phi_{\mu\bar{\mu}}(x',\bar{x}') = \cos\frac{(\mu-\bar{\mu})\pi}{2}\,\delta(x'-\bar{x}')$$
$$+\begin{cases} \dfrac{2\Gamma[(\mu+\bar{\mu})/2+1]}{\Gamma[(\bar{\mu}-\mu)/2]\Gamma(\mu+1)}\dfrac{1}{\bar{x}'}\left(\dfrac{x'^2}{\bar{x}'^2}\right)^{\mu/2+1/4} \\ \times F\left(\dfrac{\mu+\bar{\mu}}{2}+1,\ \dfrac{\mu-\bar{\mu}}{2}+1;\ \mu+1;\ \dfrac{x'^2}{\bar{x}'^2}\right), \quad x'<\bar{x}' \\ \dfrac{2\Gamma[(\mu+\bar{\mu})/2+1]}{\Gamma[(\mu-\bar{\mu})/2]\Gamma(\bar{\mu}+1)}\dfrac{1}{x'}\left(\dfrac{\bar{x}'^2}{x'^2}\right)^{\bar{\mu}/2+1/4} \\ \times F\left(\dfrac{\mu+\bar{\mu}}{2}+1,\ \dfrac{\bar{\mu}-\mu}{2}+1,\ \bar{\mu}+1;\ \dfrac{\bar{x}'^2}{x'^2}\right), \quad \bar{x}'<x', \end{cases} \tag{5.13}$$

where Γ is a gamma and F a hypergeometric function.[20] We also introduced a notation $\phi_{\mu\bar{\mu}}(x',\bar{x}')$ for this function, as we shall need it in Sec. 7 in the discussion of problems in two dimensional configuration space. We note that when used inside of an integral with respect to $x',\bar{x}'$, the curly bracket part of $\langle x'|\bar{x}'\rangle$ should be interpreted in the sense of a principal value as indicated in equation (A11) of the Appendix.

For the present problem the unitary representation $\langle x'|\bar{x}'\rangle$ in quantum mechanics can also be obtained by solving directly Eqs. (4.5a, b), which now become

$$\left(-\frac{\partial^2}{\partial x'^2} + \frac{\lambda^2}{x'^2} + \frac{\partial^2}{\partial \bar{x}'^2} - \frac{\bar{\lambda}^2}{\bar{x}'^2}\right)\langle x'|\bar{x}'\rangle = 0, \tag{5.14a}$$

$$\left(x'\frac{\partial}{\partial x'} + \bar{x}'\frac{\partial}{\partial \bar{x}'} + 1\right)\langle x'|\bar{x}'\rangle = 0. \tag{5.14b}$$

Introducing the variables ρ, ϑ by the definitions

$$x' = \rho\cosh\vartheta,\quad \bar{x}' = \rho\sinh\vartheta, \tag{5.15}$$

Eq. (5.14b) becomes

$$\left(\rho\frac{\partial}{\partial\rho}+1\right)\langle x'|\bar{x}'\rangle = 0, \tag{5.16}$$

whose solution for $\rho\neq 0$ is

$$\langle x'|\bar{x}'\rangle = \rho^{-1} f(\vartheta). \tag{5.17}$$

Substituting in (5.14a) we obtain for $f(\vartheta)$ the ordinary differential equation

$$\left(\frac{d^2}{d\vartheta^2} + \frac{\lambda^2}{\cosh^2\vartheta} - \frac{\bar{\lambda}^2}{\sinh^2\vartheta} - 1\right) f(\vartheta) = 0. \tag{5.18}$$

Now introducing the variable

$$z = \tanh^2\vartheta = (\bar{x}'/x')^2 \tag{5.19}$$

and the new function $F(z)$ through the relation

$$f(\vartheta) = f\left(\tfrac{1}{2}\ln\frac{1+\sqrt{z}}{1-\sqrt{z}}\right) \equiv (1-z)^{1/2} z^\tau F(z), \tag{5.20}$$

where

$$\tau = \tfrac{1}{2}\bar{\mu} + \tfrac{1}{4} \tag{5.21}$$

and $\mu, \bar{\mu}$ are related by (5.6) with $\lambda, \bar{\lambda}$, we obtain for $F(z)$ the equation

$$z(1-z)\frac{d^2F}{dz^2} + [(\bar{\mu}+1) - (\bar{\mu}+3)z]\frac{dF}{dz} + \left[-\left(\frac{\bar{\mu}}{2}+1\right)^2 + \left(\frac{\mu}{2}\right)^2\right]F$$
$$= 0, \tag{5.22}$$

which is a hypergeometric equation,[20] whose solution appears in the curly bracket part of (5.13).

We have determined $\langle x'|\bar{x}'\rangle$ by two methods in a particular case of potentials of the form illustrated in Fig. 1 and whose spectra are limited by $0 \leq E < \infty$. The possibility of extending the analysis to other cases will be discussed in the following sections.

As a final point we note that we made a very particular choice for the variables $T, \bar{T}$ canonically conjugate to $H, \bar{H}$. We could have certainly added an arbitrary function $g(H)/(2H)$ of H to T and in that case Eqs. (5.4) would become

$$\tfrac{1}{2}p^2 + \frac{\lambda^2}{2x^2} = \tfrac{1}{2}\bar{p}^2 + \frac{\bar{\lambda}^2}{2\bar{x}^2}, \quad xp + g(H) = \bar{x}\bar{p}. \tag{5.23}$$

The solution is again given by (5.9), but now the equation that determines the phase $\chi(k)$ is

$$k\frac{\partial \chi}{\partial k} = g(\tfrac{1}{2}k^2). \tag{5.24}$$

Obviously for any phase $\chi(k)$ we have a canonical transformation to which it corresponds.

6. CANONICAL TRANSFORMATIONS RELATING HAMILTONIANS WITH THE SAME DISCRETE SPECTRA

In Sec. 4 we derived a procedure for determining $\langle x'|\bar{x}'\rangle$ for canonical transformations that take a Hamiltonian H into another one $\bar{H}$ that has the same continuous spectrum $0 \leq E < \infty$. It seems that the procedure is feasible for discrete spectra so long as they are the same. We shall illustrate the approach by discussing a one-dimensional problem in which we have an oscillator potential plus a centrifugal term, i.e., when the classical Hamiltonian can be written as

$$H = \tfrac{1}{2}(p^2 + \lambda^2/x^2 + x^2). \tag{6.1}$$

From (4.2) the canonical conjugate variable T associated with H takes the form

$$\begin{aligned} T &= \tfrac{1}{2}\arcsin\frac{x^2 - H}{(H^2 - \lambda^2)^{1/2}} + \frac{\pi}{4} \\ &= -\tfrac{1}{2}\arcsin[(H^2 - \lambda^2)^{-1/2}2I_1] + \frac{\pi}{4}, \end{aligned} \tag{6.2}$$

where for later convenience we add a constant $\pi/4$ to the indefinite integral in (4.2), which obviously does not alter the fact that $\{T, H\} = 1$. Furthermore, we introduce the generators of the Lie algebra of the dynamical group $O(2,1)$ of the problem whose Hamiltonian is (6.1), i.e.,[17]

$$\begin{aligned} &I_1 = \tfrac{1}{4}(p^2 + \lambda^2/x^2 - x^2), \quad I_2 = \tfrac{1}{4}(xp + px), \\ &I_3 = \tfrac{1}{4}(p^2 + \lambda^2/x^2 + x^2) \\ &\quad = \tfrac{1}{2}H. \end{aligned} \tag{6.3}$$

We are now interested in the representation $\langle x'|\bar{x}'\rangle$ in quantum mechanics of the canonical transformation that takes H into

$$\bar{H} = \tfrac{1}{2}\left(\bar{p}^2 + \frac{\bar{\lambda}^2}{\bar{x}^2} + \bar{x}^2\right), \tag{6.4}$$

where the strength of the centrifugal term $\bar{\lambda} \neq \lambda$. Considering $H, \bar{H}$ as quantum mechanical operators we can obtain[17]

$$H\left(x', -i\frac{\partial}{\partial x'}\right)\psi_n(x') = (2n + \mu + 1)\psi_n(x'), \tag{6.5a}$$

$$\bar{H}\left(\bar{x}', -i\frac{\partial}{\partial \bar{x}'}\right)\bar{\psi}_n(\bar{x}') = (2n + \bar{\mu} + 1)\bar{\psi}_n(\bar{x}'), \tag{6.5b}$$

where $\mu, \bar{\mu}$ are related with $\lambda, \bar{\lambda}$ by (5.6) and the $\psi_n(x')$ are normalized, i.e.,

$$\int \psi_{n'}^*(x')\psi_n(x')\,dx' = \delta_{n'n}, \tag{6.6}$$

and similarly for $\bar{\psi}_n(\bar{x}')$. Furthermore, considering the generators I_i $(i = 1, 2, 3)$ of (6.3) of the dynamical group $O(2,1)$ as quantum mechanical operators, we have the relations[17]

$$I_\pm\psi_n(x') = [(n + \mu + \tfrac{1}{2} \pm \tfrac{1}{2})(n + \tfrac{1}{2} \pm \tfrac{1}{2})]^{1/2}\psi_{n\pm1}(x'), \tag{6.7a}$$

$$\bar{I}_\pm\bar{\psi}_n(\bar{x}') = [(n + \bar{\mu} + \tfrac{1}{2} \pm \tfrac{1}{2})(n + \tfrac{1}{2} \pm \tfrac{1}{2})]^{1/2}\bar{\psi}_{n\pm1}(\bar{x}'), \tag{6.7b}$$

where $I_\pm = I_1 \pm iI_2$ and the $\bar{I}_i$ $(i = 1, 2, 3)$ are defined as in (6.3) with all variables and parameters being replaced by the barred ones. We note furthermore that from the relation between I_3 and H in (6.3) and Eqs. (6.5), (6.7) we have that the operators η, ξ determined by

$$\eta \equiv [I_3 + (\mu - 1)/2]^{-1/2}I_+, \tag{6.8a}$$

$$\xi \equiv \eta^\dagger = I_-[I_3 + (\mu - 1)/2]^{-1/2}, \tag{6.8b}$$

which are well defined when operating on $\psi_n(x')$, give

$$\eta\psi_n(x') = \sqrt{n+1}\,\psi_{n+1}(x'), \tag{6.9a}$$

$$\xi\psi_n(x') = \sqrt{n}\,\psi_{n-1}(x'), \tag{6.9b}$$

i.e., they act as creation and annihilation operators of an ordinary oscillator. A similar relation to (6.9) holds for $\bar{\eta}, \bar{\xi}$ determined by (6.8) when we replace I_i by $\bar{I}_i$, and μ by $\bar{\mu}$.

We note that from (6.5) the operators

$$\tfrac{1}{2}(H - \mu - 1), \quad \tfrac{1}{2}(\bar{H} - \bar{\mu} - 1) \tag{6.10}$$

have the same discrete spectrum n which, from (6.9), implies that they can be written as

$$\eta\xi = \tfrac{1}{2}(H - \mu - 1), \quad \bar{\eta}\,\bar{\xi} = \tfrac{1}{2}(\bar{H} - \bar{\mu} - 1). \tag{6.11}$$

We can also check (6.10) directly by using (6.8) and the fact that the quantum mechanical Casimir operator[14,17] is

$$I_+I_- - I_3(I_3 - 1) = -\tfrac{1}{4}(\lambda^2 - \tfrac{3}{4}) = \tfrac{1}{4}(1 - \mu^2). \tag{6.12}$$

We wish now to establish the equations giving the canonical transformation that takes the Hamiltonian $H - \mu - 1$ into $\bar{H} - \bar{\mu} - 1$. From (4.4a,b) we see that these equations could be written as

$$H - \mu - 1 = \bar{H} - \bar{\mu} - 1, \tag{6.13a}$$

$$T = \bar{T}, \tag{6.13b}$$

where we used the definitions (6.1) for H, (6.4) for $\bar{H}$, (6.2) for T, and a similar one for $\bar{T}$. Equations (6.13) prove inconvenient when we go to the quantum picture and we prefer to substitute them by

$$\eta = \bar{\eta}, \tag{6.14a}$$

$$\xi = \bar{\xi}, \tag{6.14b}$$

which we will show to be equivalent. In (6.14) η, ξ are defined by (6.8) where the I_i $(i = 1, 2, 3)$ are given by

(6.3) with x, p having the classical meaning. The definition of $\bar{\eta}, \bar{\xi}$ is similar in terms of $\bar{x}, \bar{p}$ and with $\bar{\lambda}$ replacing λ. Note that here η, ξ, $\bar{\eta}, \bar{\xi}$ are understood as classical observables and thus the order of the function of x, p, $\bar{x}, \bar{p}$ appearing in their definitions is irrelevant. Using (6.11) we see that (6.13a) follows immediately from (6.14). To obtain (6.13b) we consider the angle variables[19,21]

$$w \equiv -\frac{i}{2}\ln\frac{\eta}{\xi} = -\frac{i}{2}\ln\frac{I_+}{I_-} = -i\ln\frac{I_1+iI_2}{I_1-iI_2} = \arctan\frac{I_2}{I_1}$$
$$= \arccos\frac{I_1}{\sqrt{I_1^2+I_2}} = \arccos\frac{2I_1}{\sqrt{H^2-\lambda^2}}$$
$$= -\arcsin\frac{2I_1}{\sqrt{H^2-\lambda^2}} + \frac{\pi}{2}, \tag{6.15}$$

which is canonically conjugate[19,21] to the action variable $J=\eta\xi$. In (6.15) we made use of the relation $H=2I_3$ and the *classical* Casimir operator relation

$$I_1^2 + I_2^2 - I_3^2 = -\tfrac{1}{4}\lambda^2. \tag{6.16}$$

From (6.14) we then arrive at equivalent equations for the canonical transformation of the type

$$w = \bar{w}, \tag{6.17a}$$

$$J = \bar{J}, \tag{6.17b}$$

where w, J are given in the previous paragraph and $\bar{w}, \bar{J}$ can be defined in a similar way in terms of the barred quantities. But from (6.11), (6.15) we see that Eqs. (6.13) and (6.17) are equivalent and thus we may use (6.14) as the definition of the classical canonical transformation.

To determine the unitary representation in quantum mechanics of the classical canonical transformation defined by (6.14), we return to Eq. (2.5) where we identify

$$\eta = f_1(x,p), \quad \xi = f_2(x,p),$$

$$\bar{\eta} = \bar{f}_1(\bar{x},\bar{p}), \quad \bar{\xi} = \bar{f}_2(\bar{x},\bar{p}).$$

Equations (2.24) that determine $\langle x'|\bar{x}'\rangle$ then become

$$\eta\langle x'|\bar{x}'\rangle = (\bar{\eta}^\dagger)^*\langle x'|\bar{x}'\rangle = \bar{\xi}^*\langle x'|\bar{x}'\rangle, \tag{6.18a}$$

$$\xi\langle x'|\bar{x}'\rangle = (\bar{\xi}^\dagger)^*\langle x'|\bar{x}'\rangle = \bar{\eta}^*\langle x'|\bar{x}'\rangle, \tag{6.18b}$$

where we made use of the fact that in the quantum mechanical picture the definitions (6.8) imply

$$\eta^\dagger = \xi, \quad \xi^\dagger = \eta, \quad \bar{\eta}^\dagger = \bar{\xi}, \quad \bar{\xi}^\dagger = \bar{\eta}. \tag{6.19}$$

We note also that from the definitions (6.8) η, ξ, $\bar{\eta}$, $\bar{\xi}$ are *real* differential operators, i.e., $\eta=\eta^*$, $\xi=\xi^*$. To solve Eqs. (6.18) we propose a development of $\langle x'|\bar{x}'\rangle$ in terms of the complete set of eigenfunctions (6.5a,b), i.e.,

$$\langle x'|\bar{x}'\rangle = \sum_{n,\bar{n}=0}^{\infty} a_{n\bar{n}}\,\psi_n(x')\,\bar{\psi}_{\bar{n}}^*(\bar{x}') \tag{6.20}$$

where the $a_{n\bar{n}}$ are so far undetermined coefficients. Using (6.9) and similar ones for $\bar{\eta}$, $\bar{\xi}$, Eqs. (6.18a,b) lead immediately to

$$\sum_{n,\bar{n}} \sqrt{n+1}\; a_{n\bar{n}}\,\psi_{n+1}(x')\,\bar{\psi}_{\bar{n}}^*(\bar{x}')$$
$$= \sum_{n,\bar{n}} \sqrt{n}\; a_{n\bar{n}}\,\psi_n(x')\,\bar{\psi}_{\bar{n}-1}^*(\bar{x}'), \tag{6.21a}$$

$$\sum_{n,\bar{n}} \sqrt{n}\; a_{n\bar{n}}\,\psi_{n-1}(x')\,\bar{\psi}_{\bar{n}}^*(\bar{x}')$$
$$= \sum_{n,\bar{n}} \sqrt{\bar{n}+1}\; a_{n\bar{n}}\,\psi_n(x')\,\bar{\psi}_{\bar{n}+1}^*(\bar{x}') \tag{6.21b}$$

which imply that, when in the left-hand side of (6.21a) and the right-hand side of (6.21b) we change n by $n-1$, $\bar{n}$ by $\bar{n}-1$, we obtain

$$\sqrt{n}\; a_{n-1,\bar{n}-1} = \sqrt{\bar{n}}\; a_{n\bar{n}}, \tag{6.22a}$$

$$\sqrt{n}\; a_{n\bar{n}} = \sqrt{\bar{n}}\; a_{n-1,\bar{n}-1} \tag{6.22b}$$

which give

$$(n-\bar{n})\,a_{n\bar{n}} = 0 \tag{6.22c}$$

or $a_{n\bar{n}}=0$ if $n\neq\bar{n}$. Furthermore,

$$a_{n-1,n-1} = a_{nn}; \tag{6.22d}$$

so, finally, we get

$$a_{n\bar{n}} = a_{00}\,\delta_{n\bar{n}} \tag{6.22e}$$

and the a_{00} has to be 1, or at most a constant phase, for the normalization condition (4.5c) to be satisfied. Thus we have

$$\langle x'|\bar{x}'\rangle = \sum_{n=0}^{\infty} \psi_n(x')\,\bar{\psi}_n^*(\bar{x}'). \tag{6.23}$$

We have determined the unitary representation of the classical canonical transformation that maps $H-\mu-1$ into $\bar{H}-\bar{\mu}-1$. By analogy with the discussion in Sec. 4, i.e., formula (4.7), we could have expected $\langle x'|\bar{x}'\rangle$ to be given by (6.23) with some phase factor $\exp(i\chi_n)$. The present discussion shows that this phase factor is $\chi_n=0$ if our wavefunctions are defined by (6.5)–(6.7).

Having analyzed problems in one-dimensional configuration spaces having continuous and discrete spectra, we turn our attention to problems in higher-dimensional configuration spaces.

7. CANONICAL TRANSFORMATIONS RELATING HAMILTONIANS IN HIGHER-DIMENSIONAL CONFIGURATION SPACES

So far we have discussed the representation of canonical transformations that take a Hamiltonian H in a one-dimensional configuration space into an $\bar{H}$ with the same spectrum. When we have more than one dimension we wish to find the canonical transformation, as well as its representation in quantum mechanics, that takes not only the Hamiltonian but also the other integrals of motion of one problem into those of another. Again we shall restrict ourselves to a specific example, in this case in two-dimensional configuration space, to outline as clearly as possible what seem to us the essential ideas required in this case.

We denote by x_i, p_i and $\bar{x}_i, \bar{p}_i$ $(i=1,2)$ the coordinates and momenta in our two phase spaces, and introduce the notations

$$x_\pm = \frac{1}{\sqrt{2}}(x_1 \pm i x_2), \quad p_\pm = \frac{1}{\sqrt{2}}(p_1 \mp i p_2), \tag{7.1}$$

where $x_\pm$, $p_\pm$ are canonically conjugate. Furthermore,

we have also, in the classical picture, the polar variables

$$x_\pm = \frac{1}{\sqrt{2}} r \exp(\pm i\varphi), \quad p_\pm = \frac{1}{\sqrt{2}} \exp(\mp i\varphi) \left(p_r \mp \frac{i}{r} p_\varphi \right) \tag{7.2}$$

for which

$$\{r, p_\varphi\} = \{\varphi, p_r\} = \{r, \varphi\} = \{p_r, p_\varphi\} = 0,$$
$$\{r, p_r\} = \{\varphi, p_\varphi\} = 1. \tag{7.3}$$

Similar relations hold for the barred variables $\bar{x}_i$, $\bar{p}_i$ $(i = 1, 2)$.

In the x_i, p_i classical phase space we take now the Hamiltonian H and the extra integral of motion M as those corresponding to the free particle in the plane, i.e.,

$$H = \tfrac{1}{2}(p_1^2 + p_2^2) = \tfrac{1}{2}(p_r^2 + p_\varphi^2/r^2), \tag{7.4a}$$

$$M = x_1 p_2 - x_2 p_1 = p_\varphi. \tag{7.4b}$$

In the barred variables we take $\bar{H}, \bar{M}$ defined by

$$\bar{H} = \tfrac{1}{2}\bar{p}_r^2, \tag{7.5a}$$

$$\bar{M} = \bar{p}_\varphi. \tag{7.5b}$$

Our fist question concerns the canonical conjugate variables to H, M, which we denote respectively as T, N, as well as the $\bar{T}$, $\bar{N}$ associated with $\bar{H}, \bar{M}$. They of course satisfy

$$\{H, M\} = \{T, N\} = \{H, N\} = \{M, T\} = 0,$$
$$\{T, H\} = \{N, M\} = 1 \tag{7.6}$$

and similarly for the barred quantities.

From (7.3), (7.5a, b) we immediately see that

$$\bar{T} = \bar{r}/\bar{p}_r = \bar{r}\bar{p}_r/\bar{p}_r^2, \tag{7.5c}$$

$$\bar{N} = \bar{\varphi}, \tag{7.5d}$$

but the answer is not as trivial for T, N. From (7.4) we have

$$H = p_+ p_-, \quad M = i(x_+ p_+ - x_- p_-) \tag{7.7a}$$

which satisfy $\{H, M\} = 0$, as the Poisson bracket of two observables in terms of $x_\pm$, $p_\pm$ is

$$\{F, G\} = \left(\frac{\partial F}{\partial x_+} \frac{\partial G}{\partial p_+} - \frac{\partial F}{\partial p_+} \frac{\partial G}{\partial x_+} \right) + \left(\frac{\partial F}{\partial x_-} \frac{\partial G}{\partial p_-} - \frac{\partial F}{\partial p_-} \frac{\partial G}{\partial x_-} \right). \tag{7.7b}$$

As in the one-dimensional case we can still think of T as the time when we solve the equations of motion. As p_φ is an integral of motion, the Hamiltonian (7.4a) is actually one with a centrifugal force and, as we showed in Sec. 5, the T can be written as

$$T = \frac{r p_r}{2H} = \frac{x_+ p_+ + x_- p_-}{2 p_+ p_-} = \frac{1}{2}\left(\frac{x_+}{p_-} + \frac{x_-}{p_+} \right). \tag{7.4c}$$

We immediately check, using either (7.3) or (7.7b), that $\{T, H\} = 1$, $\{T, M\} = 0$. We must now determine N through the remaining equations in (7.6), i.e.,

$$\{H, N\} = -\left(p_- \frac{\partial N}{\partial x_+} + p_+ \frac{\partial N}{\partial x_-} \right) = 0, \tag{7.8a}$$

$$\{N, M\} = i\left(x_+ \frac{\partial N}{\partial x_+} - p_+ \frac{\partial N}{\partial p_+} - x_- \frac{\partial N}{\partial x_-} + p_- \frac{\partial N}{\partial p_-} \right) = 1, \tag{7.8b}$$

$$\{T, N\} = \tfrac{1}{2}(p_+ p_-)^{-1} \{x_+ p_+ + x_- p_-, N\}$$
$$= \tfrac{1}{2}(p_+ p_-)^{-1}\left(p_+ \frac{\partial N}{\partial p_+} - x_+ \frac{\partial N}{\partial x_+} + p_- \frac{\partial N}{\partial p_-} - x_- \frac{\partial N}{\partial x_-} \right) = 0, \tag{7.8c}$$

where in (7.8c) we already made use of (7.8a). From (7.8b, c) the solution

$$N = -i(\alpha \ln x_+ - \beta \ln p_+ - \gamma \ln x_- + \delta \ln p_-) \tag{7.9a}$$

is suggested, in which we must have

$$\alpha + \beta + \gamma + \delta = 1, \quad -\alpha - \beta + \gamma + \delta = 0. \tag{7.9b}$$

Substituting this N in (7.8a) we obtain furthermore

$$\alpha(p_-/x_+) - \gamma(p_+/x_-) = 0. \tag{7.9c}$$

From (7.9) we then get

$$\alpha = \gamma = 0, \quad \beta = \delta = \tfrac{1}{2} \tag{7.10}$$

and thus, using (7.2),

$$N = \frac{i}{2} \ln \frac{p_+}{p_-} = \varphi + \arctan \frac{p_\varphi}{r p_r}. \tag{7.4d}$$

A canonical transformation that maps H, M onto $\bar{H}, \bar{M}$ is then given by the equations

$$H = \bar{H}, \quad M = \bar{M}, \quad T = \bar{T}, \quad N = \bar{N}. \tag{7.11}$$

As indicated in the previous sections, it may prove more convenient, for later translation to quantum mechanics, to use instead of the equations (7.11) some functions of the observables involved. Specifically, the classical canonically transformation mapping H, M onto $\bar{H}, \bar{M}$ can equally well be defined by the equations

$$H = \bar{H}, \quad M = \bar{M}, \quad 2HT = 2\bar{H}\bar{T}, \quad \exp(iN) = \exp(i\bar{N}) \tag{7.12}$$

and it will be these ones that we shall use.

The question now is how to obtain the representation in quantum mechanics associated with the classical canonical transformation determined by Eqs. (7.12). Using a notation similar to that of the previous sections, we shall denote this representation by

$$\langle r'\varphi' | r'\bar{\varphi}' \rangle. \tag{7.13}$$

The equations satisfied by $\langle r'\varphi' | \bar{r}'\bar{\varphi}' \rangle$ are of the type (2.24), but before deriving them from (7.12), we shall express all the operators associated with the classical observables appearing there in a Hermitian form. As the volume element in configuration space is $r\,dr\,d\varphi$, the operators p_r, p_φ to be Hermitian must take the form[22]

$$p_r = -i \frac{1}{\sqrt{r}} \frac{\partial}{\partial r} \sqrt{r}, \tag{7.13'}$$

$$p_\varphi = -i \frac{\partial}{\partial \varphi}. \tag{7.13''}$$

The same applies to $\bar{p}_r, \bar{p}_\varphi$ canonically conjugate to

$\bar{r}, \bar{\varphi}$. The Hermitian form of the H of (7.4a) is

$$H=\frac{1}{2}\left(-\frac{1}{r}\frac{\partial}{\partial r}\,r\frac{\partial}{\partial r}-\frac{1}{r^2}\frac{\partial^2}{\partial\varphi^2}\right) \tag{7.14a}$$

as it is related to the Hamiltonian of the free particle in the plane. From (7.13′) we have for $\bar{H}$ of (7.5a) the Hermitian operator

$$\bar{H}=-\frac{1}{2}\frac{1}{\sqrt{\bar{r}}}\frac{\partial^2}{\partial\bar{r}^2}\sqrt{\bar{r}}\,. \tag{7.14b}$$

We now turn our attention to the classical observables

$$2HT=rp_r,\quad 2\bar{H}\,\bar{T}=\bar{r}\bar{p}_{\bar{r}}. \tag{7.15}$$

From the form $r\,dr\,d\varphi$ of the volume element, the corresponding Hermitian operators are

$$2HT=-i\frac{\partial}{\partial r}r, \tag{7.16a}$$

$$2\bar{H}\,\bar{T}=-i\frac{\partial}{\partial\bar{r}}\,\bar{r}. \tag{7.16b}$$

Finally, we look into

$$\exp(iN)=\left(\frac{p_-}{p_+}\right)^{1/2}=\frac{p_-}{H^{1/2}}, \tag{7.17a}$$

$$\exp(i\bar{N})=\exp(i\bar{\varphi}), \tag{7.17b}$$

where as $H=p_+p_-/2$ commutes with p_-, the order in (7.17a) is irrelevant when we pass to quantum mechanics. We note that

$$p_-=-i\frac{\partial}{\partial x_-'}=\frac{1}{\sqrt{2}}\exp(i\varphi')\left(\frac{1}{i}\frac{\partial}{\partial r'}+\frac{1}{r'}\frac{\partial}{\partial\varphi'}\right) \tag{7.18}$$

has the hermiticity property

$$p_-^\dagger=p_+. \tag{7.19}$$

This, together with (7.17) implies,

$$[\exp(iN)]^\dagger=\exp(-iN), \tag{7.20a}$$

$$[\exp(i\bar{N})]^\dagger=\exp(-i\bar{N}) \tag{7.20b}$$

and thus $N, \bar{N}$ are already Hermitian.

Equations (2.24) in which we take

$$f_1=H,\ f_2=M,\ f_3=2HT,\ f_4=\exp(iN), \tag{7.21}$$

and similarly for the barred $\bar{f}_\alpha(\alpha=1,2,3,4)$, give then

$$-\frac{1}{2}\left(\frac{1}{r'}\frac{\partial}{\partial r'}\,r'\frac{\partial}{\partial r'}+\frac{\partial^2}{\partial\bar{r}'^2}\frac{\partial^2}{\partial\bar{\varphi}'^2}\right)\langle r'\varphi'|\bar{r}'\bar{\varphi}'\rangle$$

$$=-\frac{1}{2}\frac{1}{\sqrt{\bar{r}'}}\frac{\partial^2}{\partial\bar{x}'^2}\sqrt{\bar{r}'}\,\langle r'\varphi'|\bar{r}',\bar{\varphi}'\rangle, \tag{7.22a}$$

$$\frac{1}{i}\frac{\partial}{\partial\varphi'}\langle r'\varphi'|\bar{r}'\bar{\varphi}'\rangle=-\frac{1}{i}\frac{\partial}{\partial\bar{\varphi}'}\langle r'\varphi'|\bar{r}'\bar{\varphi}'\rangle, \tag{7.22b}$$

$$\frac{1}{i}\frac{\partial}{\partial r'}\,r'\langle r'\varphi'|\bar{r}'\bar{\varphi}'\rangle=-\frac{1}{i}\frac{\partial}{\partial\bar{r}'}\,\bar{r}'\langle r'\varphi'|\bar{r}'\bar{\varphi}'\rangle, \tag{7.22c}$$

$$H^{-1/2}\frac{1}{\sqrt{2}}\exp(i\varphi')\left(\frac{1}{i}\frac{\partial}{\partial r'}+\frac{1}{r'}\frac{\partial}{\partial\varphi'}\right)\langle r'\varphi'|\bar{r}'\bar{\varphi}'\rangle$$

$$=\exp(-i\bar{\varphi}')\,\langle r'\varphi'\rangle\,\langle\bar{r}'\bar{\varphi}'\rangle. \tag{7.22d}$$

From (7.22a, b) we can write

$$\langle r'\varphi'|\bar{r}'\bar{\varphi}'\rangle$$

$$=\sum_{m=-\infty}^{\infty}\frac{1}{2\pi}\int_0^\infty\exp[i\chi_m(k)][i^mJ_m(kr')\exp(im\varphi')\,J_{1/2}(k\bar{r}')$$

$$\times\exp(-im\bar{\varphi}')]\,k\,dk, \tag{7.23}$$

where we have only a phase factor $\exp[i\chi_m(k)]$, as the eigenfunctions of H, M are already normalized. Now applying Eq. (7.22c), we have

$$\left(r'\frac{\partial}{\partial r'}+\bar{r}'\frac{\partial}{\partial\bar{r}'}+2\right)\langle r'\varphi'|\bar{r}'\bar{\varphi}'\rangle$$

$$=\sum_m\frac{i^m}{2\pi}\int_0^\infty\exp[i\chi_m(k)]\left(k\frac{\partial}{\partial k}+2\right)[J_m(kr')$$

$$\times\exp(im\varphi')\,J_{1/2}(k\bar{r}')\exp(-im\bar{\varphi}')]\,k\,dk$$

$$=-\sum_m\frac{i^{m+1}}{2\pi}\int_0^\infty\frac{\partial\chi_m}{\partial k}\exp[i\chi_m(k)][J_m(kr')$$

$$\times\exp(im\varphi')\,J_{1/2}(k\bar{r}')\exp(-im\bar{\varphi}')]\,k^2\,dk=0, \tag{7.24}$$

where as the functions in the square bracket depend only on $kr', k\bar{r}'$, we can replace $r'\partial/\partial r'+\bar{r}'\partial/\partial\bar{r}'$ by $k\partial/\partial k$. The last integral comes from integration by parts.

We now apply (7.22d), remembering that[23]

$$p_-i^mJ_m(kr')\exp(im\varphi')=\frac{k}{\sqrt{2}}\,i^{m+1}J_{m+1}(kr')\exp[i(m+1)\varphi'], \tag{7.25}$$

to obtain

$$\sum_{m=-\infty}^{\infty}\frac{1}{2\pi}\int_0^\infty\exp[i\chi_m(k)\,i^{m+1}J_{m+1}(kr')$$

$$\times\exp[i(m+1)\varphi']\,J_{1/2}(k\bar{r}')\exp(-im\bar{\varphi}')\,k\,dk$$

$$=\sum_{m=-\infty}^{\infty}\frac{1}{2\pi}\int_0^\infty\exp[i\chi_m(k)]\,i^mJ_m(kr')$$

$$\times\exp(im\varphi')\,J_{1/2}(k\bar{r}')\exp[-i(m-1)\bar{\varphi}']\,k\,dk. \tag{7.26}$$

Changing m into $m+1$ in the right-hand side of (7.26), we obtain from this equation and (7.24) that

$$\chi_{m+1}(k)=\chi_m(k),\quad \chi_m(k)\ \text{independent of } k. \tag{7.27}$$

Thus $\chi_m(k)$ is independent of m and k and as the representation is anyway determined only up to a phase we can take $\chi_m(k)=0$ in (7.23). The integration with respect to k is then identical to (4.13) except for a factor $(r'\bar{r}')^{1/2}$ and thus we have

$$\langle r'\varphi'|\bar{r}'\bar{\varphi}'\rangle$$

$$=\sum_{m=-\infty}^{\infty}\frac{i^m}{2\pi}\,(r'\bar{r}')^{-1/2}\,\phi_{m,1/2}(r',\bar{r}')\exp(im)(\varphi'-\bar{\varphi}'). \tag{7.28}$$

We have shown the possible procedure to be followed in the determination of the representations in quantum mechanics of canonical transformations in higher-dimensional phase space. In the last section we now summarize the conclusions that follow from the specific examples discussed here, as well as indicate possible procedures when the spectra of the operators, transformed into each other by the canonical transformations, are not the same.

8. CONCLUSIONS

We have analyzed some examples of representations in quantum mechanics of nonlinear canonical transformations, in which difficulties number one and two of the Introduction (i.e., fractional or negative powers of the differential operators and order of the factors when we replace classical observables by noncommuting operators in the equations that determine the representation) can be overcome by dealing with suitable functions f_α of the coordinates and momenta, instead of using the latter directly. In the examples studied, the relation between the old x_i, p_i and the new $\bar{x}_i$, $\bar{p}_i$ was rather complicated, while the functions f_α were simple enough to permit their quantization without ambiguity.

Of course one cannot tell whether or not it will be possible to find such suitable combinations for an arbitrary problem. There are, however, ways to deal with the problem of the quantization of arbitrary classical quantities in a consistent manner. One such formalism is that provided by the Weyl–Moyal transform.[10,16] It might be interesting to explore the possibility of stating the general problem in this language.

Concerning difficulty number three of the Introduction, i.e., when the new and old observables have different spectra in the quantum mechanical picture, it has not been examined at all in the present paper. As it was mentioned in Sec. 2, an appropriate starting point might be provided by relation (2.16) for U instead of Eq. (2.15). Now, however, one does not ask U to be unitary; it might not even have an inverse. If one is dealing with canonical transformations that depend on a set of parameters λ, then the operators $U(\lambda)$ will follow the same multiplication law as the corresponding classical transformations, so that they constitute a representation of the latter in the usual group-theoretical sense. However, the absence of an inverse indicates that we are now dealing with a semigroup rather than with a group of transformations.

The approach mentioned in the last paragraphs could be applied to some simple canonical transformations that have the feature of modifying the spectra of the quantum mechanical operators. Specifically, we have in mind the canonical transformation that maps an oscillator of frequency k^{-1} (k, integer) onto another of frequency unity. We plan to discuss in future publications these types of canonical transformations and their representation in quantum mechanics.

As a final point we would like to indicate one possible application of the representations of nonlinear canonical transformations. In some cases it may be easier to solve the equations that determine the representation of a canonical transformation connecting two Hamiltonians H and $\bar{H}$, for example Eq. (5.14), than to find the eigenstates of H itself. This would then provide a procedure of determining the latter. In the case of linear canonical transformations this approach was applied[25] to give a simple solution for the problem of a particle in a constant magnetic field.

APPENDIX: PROOF OF EQ. (5.13)

From (5.9) and the result (5.12) we can write the transformation bracket $\langle x'|\bar{x}'\rangle$ as

$$\langle x'|\bar{x}'\rangle = \sqrt{x'\bar{x}'}\int_0^\infty k J_\mu(kx') J_{\bar{\mu}}(k\bar{x}')\,dk. \tag{A1}$$

This integral can be found in Ref. 20, Sec. 6.574 to be

$$\langle x'|x'\rangle = \begin{cases} 2\dfrac{\Gamma((\mu+\bar{\mu})/2+1)}{\Gamma((\bar{\mu}-\mu)/2)\,\Gamma(\mu+1)}\left(\dfrac{x'^2}{\bar{x}'^2}\right)^{\mu/2+1/4} \\ \quad\times F\left(\dfrac{\mu+\bar{\mu}}{2}+1,\ \dfrac{\mu-\bar{\mu}}{2}+1;\mu+1;\dfrac{x'^2}{\bar{x}'^2}\right), \\ \qquad\qquad (x'<\bar{x}') \qquad\qquad (A2a) \\ 2\dfrac{\Gamma((\mu+\bar{\mu})/2+1)}{\Gamma((\mu-\bar{\mu})/2)\,\Gamma(\bar{\mu}+1)}\left(\dfrac{\bar{x}'^2}{x'^2}\right)^{\bar{\mu}/2+1/4} \\ \quad\times F\left(\dfrac{\mu+\bar{\mu}}{2}+1,\ \dfrac{\bar{\mu}-\mu}{2}+1;\bar{\mu}+1;\dfrac{\bar{x}'^2}{x'^2}\right), \\ \qquad\qquad (\bar{x}'<x'). \qquad\qquad (A2b) \end{cases}$$

These results are valid as far as $x' \neq \bar{x}'$ and they diverge at $x' = \bar{x}'$. To find out the nature of the singularity, we use the form of the hypergeometric function near the value 1 of the argument (Ref. 24, Eq. 15.3.12) to write

$$\langle x'|\bar{x}'\rangle = \frac{1}{\pi}\sin\frac{\pi(\bar{\mu}-\mu)}{2}\left(\frac{1}{\bar{x}'-x'} - \frac{\bar{\mu}^2-\mu^2}{2x'}\ln|\bar{x}'-x'|\right) + \text{analytic function}. \tag{A3}$$

The integral (A1) will also contain a δ-function singularity which does not appear in (A3). The behavior of the Bessel functions at infinity is the one responsible for the singularities at $x' = \bar{x}'$. If we use the asymptotic expansion of the Bessel functions (Ref. 24, Eq. 9.2.5), the integration of the first few terms should reproduce the result (A3) plus a δ-function. This expansion yields

$$\begin{aligned}\langle x'|\bar{x}'\rangle = \frac{2}{\pi}\int_0^\infty \Bigg[&\cos\left(kx' - \frac{\mu\pi}{2} - \frac{\pi}{4}\right)\cos\left(k\bar{x}' - \frac{\bar{\mu}\pi}{2} - \frac{\pi}{4}\right) \\ &- \frac{4\bar{\mu}^2-1}{8\bar{x}'}\frac{1}{k}\cos\left(kx' - \frac{\mu\pi}{2} - \frac{\pi}{4}\right)\sin\left(k\bar{x}' - \frac{\bar{\mu}\pi}{2} - \frac{\pi}{4}\right) \\ &- \frac{4\mu^2-1}{8x'}\frac{1}{k}\cos\left(k\bar{x}' - \frac{\bar{\mu}\pi}{2} - \frac{\pi}{4}\right)\sin\left(kx' - \frac{\mu\pi}{2} - \frac{\pi}{4}\right)\Bigg]dk \\ &+ \text{analytic function}. \end{aligned} \tag{A4}$$

Using the result

$$\int_0^\infty \exp(ikx)\,dk = \pi\delta(x) + \mathcal{P}/x, \tag{A5}$$

and the fact that $x, x' > 0$, we can write

$$\begin{aligned}\int_0^\infty &\cos(kx'-\alpha)\cos(k\bar{x}'-\beta)\,dk \\ &= \frac{\pi}{2}\cos(\alpha-\beta)\,\delta(x'-\bar{x}') + \tfrac{1}{2}\sin(\alpha-\beta)\,\frac{\mathcal{P}}{x'-\bar{x}'} \\ &\quad + \text{analytic function}. \end{aligned} \tag{A6}$$

Concerning the second and third terms of (A4), the divergence at $k=0$ is only apparent, since there the series expansion, valid for large values of the argument, breaks down. The divergence at $x' = \bar{x}'$ we are

looking for can be extracted if we integrate from, say, k_0, to ∞. We use the result

$$\int_{k_0}^{\infty} \frac{\exp(ik\xi)}{ik}\, dk = \frac{\pi}{2} - \mathrm{Si}(k_0\xi) + i\mathrm{Ci}(k_0\xi), \tag{A7a}$$

$$\int_{k_0}^{\infty} \frac{\exp(-ik\xi)}{ik}\, dk = -\frac{\pi}{2} + \mathrm{Si}(k_0\xi) + i\mathrm{Ci}(k_0\xi), \tag{A7b}$$

where the cosine integral Ci and the sine integral Si are defined in Ref. 24, Eqs. 5.2.26, 5.2.27; they have, near the origin, the series expansion

$$\mathrm{Si}(z) = z + \cdots,$$
$$\mathrm{Ci}(z) = \gamma + \ln z - \frac{z^2}{4} + \cdots, \tag{A8}$$

so that

$$\int_{k_0}^{\infty} \frac{\exp(ik\xi)}{ik}\, dk = i\ln|\xi| + \text{analytic function}. \tag{A9}$$

We can therefore write

$$\int_{k_0}^{\infty} \frac{1}{k} \sin(kx' - \alpha)\cos(k\bar{x}' - \beta)\, dk$$
$$= \tfrac{1}{2}\sin(\alpha - \beta)\ln|x' - \bar{x}'| + \text{analytic function}. \tag{A10}$$

Using the results (A6) and (A10), we can then write (A4) as

$$\langle x'|\bar{x}'\rangle$$
$$= \cos\frac{(\mu - \bar{\mu})\pi}{2}\,\delta(x' - \bar{x}') + \frac{1}{\pi}\sin\frac{\pi(\bar{\mu} - \mu)}{2}\left[\frac{P}{\bar{x}' - x'} - \frac{\bar{\mu}^2 - \mu^2}{2x'}\ln|\bar{x}' - x'|\right] + \text{analytic function}, \tag{A11}$$

so that result (5.13) of the text follows.

*Member of the Instituto Nacional de Energía Nuclear.
†Member of El Colegio Nacional.

[1] M. Born, W. Heisenberg, and P. Jordan, Z. Phys. **35**, 557 (1926).
[2] P.A.M. Dirac, *Quantum Mechanics* (Clarendon, Oxford, 1958), 4th ed., pp. 103–107.
[3] E. P. Wigner, *Group Theory* (Academic, New York, 1959).
[4] G. Racah, *Group Theory and Spectroscopy*, Ergebnisse der Exakten Wissenschaften, Vol. 37 (Springer, Berlin, 1965).
[5] V. Fock, Z. Phys. **98**, 145 (1935).
[6] V. Bargmann, Z. Phys. **99**, 576 (1936).
[7] J.M. Jauch and E. L. Hill, Phys. Rev. **57**, 641 (1940).
[8] J. P. Elliot, Proc. Roy. Soc. Lond. A **245**, 128 (1958).
[9] M. Moshinsky, "Group Theory and the Few Nucleon Problem" in *Cargèse Lectures in Physics*, edited by M. Jean (Gordon and Breach, New York, 1969).
[10] L. Infeld and J. Plebański, Bull. Act. Polon. Sci. III, **3**, 95 (1955); L. Infeld and J. Plebański, Acta Phys. Polon. **14**, 41 (1955); J. Plebański, Acta Phys. Polon. **14**, 275 (1955).
[11] C. Itzykson, Commun. Math. Phys. **4**, 92 (1967); V. Bargmann, "Group Representations on Hilbert Spaces of Analytic Functions" in *Analytic Methods of Mathematical Physics*, edited by Gilbert and Newton (Gordon and Breach, New York, 1968).
[12] M. Moshinsky and C. Quesne "Oscillator Systems" in *Symmetry Properties of Nuclei*, Proceedings of the Fifteenth Solvay Conference on Physics (Gordon and Breach, London, 1974).
[13] M. Moshinsky and C. Quesne, J. Math. Phys. **12**, 1772 (1971).
[14] M. Moshinsky, SIAM J. Appl. Math. **25**, 193 (1973).
[15] P. Kramer, M. Moshinsky and T.H. Seligman, "Complex Extensions of Canonical Transformations and Quantum Mechanics" in *Group Theory and Its Applications*, edited by E.M. Loebl (Academic, New York, 1975), Vol. III.
[16] S.R. de Groot and L.G. Suttorp, *Foundations of Electrodynamics* (North-Holland, Amsterdam, 1972), Chap. VI.
[17] M. Moshinsky, T.H. Seligman and K.B. Wolf, J. Math. Phys. 13, 901 (1972).
[18] P.M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), p. 943.
[19] H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, Mass., 1959), p. 243.
[20] I. S. Gradshteyn and I. W. Ryzhik, *Tables of Integrals Series and Products* (Academic, New York, 1965).
[21] M. Moshinsky, J. Patera, and P. Winternitz, J. Math. Phys. **16**, 82 (1975).
[22] Ref. 2, pp. 152, 153.
[23] M. Moshinsky and J. Patera, J. Math. Phys. **16**, 1866 (1975).
[24] *Handbook of Mathematical Functions*, edited by M. Abramowitz and I.A. Stegun (National Bureau of Standards, Washington, D.C. 1964).
[25] M.H. Boon and T.H. Seligman, J. Math. Phys. **14**, 1224 (1973).

Higher-dimensional unifications of gravitation and gauge theories

Y. M. Cho*

The Enrico Fermi Institute and the Department of Physics, The University of Chicago, Chicago, Illinois 60637
(Received 9 April 1975)

We give a comprehensive geometric treatment of Kaluza–Klein type unifications of non-Abelian gauge theories with gravitation. The appearance of a cosmological term is noted.

1. INTRODUCTION

A five-dimensional unification of Einstein's theory of gravitation with that of an Abelian gauge field (e.g., that of electromagnetism) has been constructed a long time ago by Kaluza[1] and was further developed by Klein.[2] Kaluza's ideas have been generalized to non-abelian gauge fields by de Witt, Trautmen, and Kerner.[3] In view of the revived interest in unified field theories we wish to give here a mathematically compelling treatment of non-Abelian Kaluza-type theories.

The idea here is to introduce a $(4+n)$-dimensional ($n=$ the dimension of the internal symmetry group) Riemann space. Its quotient space by the equivalence relation of group transformations is the usual four-dimensional space–time. The enlarged space is called the principal fiber bundle space in differential geometry. Roughly speaking, the unified field theory emerges by identifying $4n$ of the new components of the metric tensor with the Yang–Mills gauge potentials of the internal symmetry group and postulating the Lagrangian of the theory to be the scalar curvature density $R_{4+n}\sqrt{-g_{4+n}}$ of the fiber bundle space. This Lagrangian equals the sum of the curvature of the four-dimensional space–time R_4, the Yang–Mills Lagrangian, and a term R_G, the curvature of the group space which plays the role of a cosmological constant.

2. THE GEOMETRY

We require the geometry of the real space–time M combined with the internal symmetry Lie group G to be that of a principal fiber bundle P with M as the base manifold and G as the structural group. Namely we want[4]:

D-1: (a) Every element a of G to induce a smooth mapping R_a of P into itself, R_a: $(p,a)\in P\times G\to pa\in P$ with $p(ab)=(pa)b$ for all a, $b\in G$, $p\in P$, and $pa=p$ for some p implying $a=e$, the unit element of the group. In short, the group G acts smoothly to the right on P without fixed point.

(b) M is the quotient space of P by the equivalence relation of the right multiplication R_a, and the projection Π: $P\to M$ is smooth.

(c) P is locally trivial, that is, for any $x\in M$ there exists an open subset U of M containing x such that the set $\Pi^{-1}(U)$ defined by $\Pi(\Pi^{-1}(U))=U$ is isomorphic to $U\times G$. By this, we mean that there exists a diffeomorphism that takes the points $p\in\Pi^{-1}(U)$ onto the points $(\Pi(p),\ \phi(p))\in U\times G$ such that $\phi(pa)=\phi(p)a$ for all $a\in G$.

Notice that the fiber over x defined as $\Pi^{-1}(x)$ is isomorphic to G (see Fig. 1). $\Pi^{-1}(x)$, a smooth submanifold itself, will be called the fiber space over x.

The above structure is natural, if one views the fiber degree of freedom $\Pi^{-1}(x)$ at each space–time point x as the local gauge degree of freedom at x, and the right multiplication R_a as a local gauge transformation corresponding to $a\in G$.

Now let us choose a coordinate basis $\xi_\mu=\partial_\mu$ ($\mu=1\cdots4$) for the base manifold M whose commutation relations are trivial:

$$[\xi_\mu,\ \xi_\nu]=[\partial_\mu,\ \partial_\nu]=0.$$

For a basis of G, choose a set of n linearly independent left invariant vector fields ξ_i $(i=4+1\cdots4+n)$ on G which are defined by

$$\forall a\in G\quad L_a\xi_i=\xi_i\quad (i=4+1\cdots4+n),$$

where $L_a\xi_i$ is the mapping of ξ_i induced by the mapping L_a of G onto itself by L_a: $b\in G\to ab\in G$ (i.e., the left multiplication).[5]

These ξ_i can also be viewed as a basis of the Lie algebra $\mathcal{G}$ of G which forms an n-dimensional vector space. The commutation relations of these vector fields are

$$[\xi_i,\ \xi_j]=f_{ij}^k\,\xi_k,$$

where the f_{ij}^k are the structure constants of the group.

Notice that this basis for G is *not* a coordinate basis[6] and the commutation relations imply that the directional derivatives of two basis vector fields do not commute in general.

Also observe that each ξ_i generates a one-parameter group of transformations $\phi_i(t)$ on G with[4]

$$\phi_i(t)\cdot a=\phi_i(t)\cdot L_a\cdot e=L_a\cdot\phi_i(t)\cdot e$$

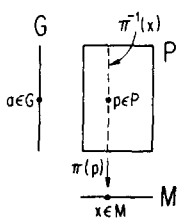


FIG. 1. The principal fiber bundle P with M as the base manifold and G as the structural group. p is a point on the fiber $\Pi^{-1}(x)$ at $x\in M$.

$$= a \cdot (\phi_i(t) \cdot e) = R_{\phi_i(t)\cdot e}\, a$$

for every $a \in G$, where the second equality comes from the left invariance of ξ_i.

The basis ξ_i $(i = 4+1 \cdots 4+n)$ of G can be mapped in a natural way into the fiber space of the bundle since $R_{\phi_i(t)\cdot e}$ can also be viewed as a one-parameter group of transformations acting *on the bundle*.

We will write the vector fields *in the bundle* induced by $R_{\phi_i(t)\cdot e}(i = 4+1 \cdots 4+n)$ as ξ_i^* $(i = 5 \cdots 4+n)$, and as usual call them the fundamental vector fields.[4] These fundamental vector fields are clearly tangent to the fiber space and form a subspace, called the vertical subspace, of the tangent space $T_p(P)$ to P at each $p \in P$.

Now, we want to give one more crucial structure, a connection Γ to our bundle. A connection Γ in P is a choice of a tangent subspace called the horizontal subspace H_p at each $p \in P$ which satisfies[4]:

D-2: (a) The tangent space T_p at each $p \in P$ is the direct sum of the vertical subspace and the horizontal subspace H_p.

(b) $\forall$ $a \in G$, $p \in P$ $H_{pa} = R_a H_p$.

(c) H_p is smooth on P.

The condition (b) implies that any two horizontal spaces at two different points on the same fiber should differ only by a gauge transformation.

Of course, on any principal fiber bundle, such a connection exists.[4]

Now, given any vector ζ at $x \in M$, we can "lift" it "horizontally" into the points $p \in \Pi^{-1}(x) \subset P$ in the following way. Consider at each point p of the fiber $\Pi^{-1}(x)$ a vector $\hat{\zeta} \in H_p$ such that $\Pi(\hat{\zeta}) = \zeta$. $\hat{\zeta}$ so defined is unique[4] and is called the horizontal lift of ζ at $p \in P$. Notice that the fiber space dependence of $\hat{\zeta}$ is given by right invariance. This is so because if $\hat{\zeta}(p)$ is the horizontal lift of ζ at p, then $R_a \hat{\zeta}(p)$ is the horizontal lift of ζ at pa. $R_a \hat{\zeta}(p)$ is horizontal and $\Pi(R_a \hat{\zeta}(p)) = \Pi(\hat{\zeta}(p)) = \zeta$. Then the uniqueness implies that $R_a \hat{\zeta}(p) = \hat{\zeta}(pa)$, that is, $\hat{\zeta}$ is invariant by R_a.[4]

In particular, the previously chosen coordinate basis $\xi_\mu = \partial_\mu$ $(\mu = 1 \cdots 4)$ of the base manifold M can be lifted horizontally into the bundle P. Let $\hat{\xi}_\mu$ $(\mu = 1 \cdots 4)$ be the horizontal lift of ξ_μ.

As for the notation, we will continue to use $i, j, k, \cdots$ $(= 4+1 \cdots 4+n)$ to label the group degree of freedom, $\mu, \nu, \alpha, \beta, \cdots$ $(= 1 \cdots 4)$, the space–time degree of freedom. For the whole bundle space degree of freedom we will use $a, b, \cdots$ $(= 1 \cdots 4+n)$.

Clearly $\hat{\xi}_\mu$ $(\mu = 1 \cdots 4)$ and ξ_i^* $(i = 4+1 \cdots 4+n)$ are $4+n$ linearly independent vector fields in the bundle space, and their commutation relations are

$$\begin{aligned} &[\xi_i^*, \xi_j^*] = f_{ij}^k \xi_k^*, \\ &[\xi_i^*, \hat{\xi}_\mu] = 0, \\ &[\hat{\xi}_\mu, \hat{\xi}_\nu] = -F_{\mu\nu}^k \xi_k^*. \end{aligned} \tag{1}$$

The first set of relations follows from the isomorphism of the ξ_i^*'s and ξ_i's. The second set is nothing more than the right invariance of the $\hat{\xi}_\mu$'s in the fiber space. To understand the third set, one has only to realize that the horizontal component of the commutator of any two horizontal lift vector fields $[\hat{\eta}, \hat{\zeta}]$ should be the same as the horizontal lift of $[\eta, \zeta]$ because

$$\Pi(H[\hat{\eta}, \hat{\zeta}]) = \Pi([\hat{\eta}, \hat{\zeta}]) = [\eta, \zeta].$$

Therefore, $[\hat{\xi}_\mu, \hat{\xi}_\nu]$ has to be vertical which is just the third equation (1). We would like to emphasize that the commutator of two horizontal vector fields *has a vertical component* in general. $F_{\mu\nu}^k$ correspond to Yang–Mills fields, which we will soon justify.

There is one more observation to make with the third commutation relations. Since the $\hat{\xi}_\mu$'s are defined to be right invariant, the commutator $-F_{\mu\nu}^k \xi_k^*$ should also be right invariant. This immediately dictates the fiber space dependence of $F_{\mu\nu}^k$. In fact, from one of the Jacobi identities

$$\begin{aligned} &[\xi_i^*, [\hat{\xi}_\mu, \hat{\xi}_\nu]] + [\hat{\xi}_\mu, [\hat{\xi}_\nu, \xi_i^*]] + [\hat{\xi}_\nu, [\xi_i^*, \hat{\xi}_\mu]] \\ &\quad = [\xi_i^*, -F_{\mu\nu}^k \xi_k^*] = -(\partial_i^* F_{\mu\nu}^k) \xi_k^* - F_{\mu\nu}^j f_{ij}^k \xi_k^* \\ &\quad = 0, \end{aligned}$$

we have

$$\partial_i^* F_{\mu\nu}^k = -f_{ij}^k F_{\mu\nu}^j. \tag{2}$$

Here the derivative $\partial_i^* = \partial_{\xi_i^*}$ is meant as the directional derivative with respect to the vector field ξ_i^*.

Of course, Eq. (2) is well known once one admits that $F_{\mu\nu}^k$ are Yang–Mills fields; for non-Abelian groups these are gauge *covariant*. They happen to be gauge invariant in the Abelian case simply because they are then neutral.

3. GAUGE TRANSFORMATIONS

So far we did not say much about $F_{\mu\nu}^k$ except that they are the vertical coefficients of the commutator $[\hat{\xi}_\mu, \hat{\xi}_\nu]$ and are gauge covariant. Now we will introduce some more concepts, those cooresponding to vector potentials in our geometry, so that one can write down $F_{\mu\nu}^k$ in terms of these vector potentials and complete their identification with Yang–Mills fields.

Given a connection Γ in P, one can define a $\mathcal{G}$-valued 1-form ω on P, i.e., a "covariant" vector field which has values not in the space of real numbers but in the vector space $\mathcal{G}$ (as before $\mathcal{G}$ is the Lie algebra of G) in the following manner.

Define ω to be a linear mapping which maps ξ_i^* into ξ_i and $\hat{\xi}_\mu$ into the zero in $\mathcal{G}$. This ω is called the connection form of the given connection Γ.

From the definition D-2, ω should satisfy[4]:

D-2′: (a) $\forall$ ξ_i^* $\omega(\xi_i^*) = \xi_i$.

(b) $\forall$ $a \in G$ $R_a \cdot \omega = \mathrm{ad}(a^{-1}) \cdot \omega$

i.e., $(R_a \cdot \omega)_p(\zeta) \equiv \omega_{pa}(R_a \cdot \zeta) = \mathrm{ad}(a^{-1}) \cdot \omega_p(\zeta)$ for any $a \in G$ and any vector field ζ in $T(P)$, where $\mathrm{ad}(a)$ is the adjoint representation of $a \in G$, a real $n \times n$ matrix representation operating on the vector $\omega(\zeta) \in \mathcal{G}$, a real $1 \times n$ column matrix.

The condition (a) is trivial by the definition. To prove (b), it is enough to verify it in the following two cases since ω is linear in ζ:

(i) If ζ is horizontal, so is $R_a \cdot \zeta$. Hence both $\omega(R_a \cdot \zeta)$ and $\mathrm{ad}(a^{-1}) \cdot \omega(\zeta)$ vanish.

(ii) If ζ is fundamental, $R_a \zeta = \mathrm{ad}(a^{-1})\zeta$ because ζ is induced by a $R_{\phi(t)\cdot e}$ and $R_a \cdot \zeta$ is by $R_a R_{\phi(t)\cdot e} R_{a^{-1}}$ $= R_{a^{-1}\cdot\phi(t)\cdot e\cdot a} = R_{a^{-1}\phi(t)a\cdot e}$ which will in turn induce $\mathrm{ad}(a^{-1}) \cdot \zeta$. Therefore $\omega_{pa}(R_a \zeta) = \mathrm{ad}(a^{-1}) \cdot \omega_p(\zeta)$.

Conversely, given any $\mathcal{G}$-valued 1-form ω which satisfies D-2′, one can always define a connection Γ whose connection form is ω. For this, one has only to define H_p as the set of tangent vectors at p which are mapped into the zero element of $\mathcal{G}$ by ω.

We can decompose ω using the basis ξ_i of $\mathcal{G}$:

$$\omega \equiv \omega^i \xi_i, \tag{3}$$

where the ω^i are now real valued 1-forms, i.e., covariant vectors in the usual sense, which satisfy

$$\begin{aligned} &\omega^i(\xi_k^*) = \omega_a^i \xi_k^{*a} = \delta_k^i, \\ &\omega^i(\hat{\xi}_\mu) = \omega_a^i \hat{\xi}_\mu^a = 0. \end{aligned} \tag{4}$$

This immediately tells us that the ω^i are nothing more than the dual vector fields to the ξ_i^*.

One can express a connection form ω on P by a family of forms each defined in an open subset U of M in the following way.[7]

Let a cross section $\sigma(x)$ be a submanifold in P which is diffeomorphic to U with $\Pi(\sigma(x)) = x$ and let $\sigma \cdot \xi_\mu$ be a vector in the tangent space $T_{\sigma(x)}(P)$ to P at $\sigma(x)$ induced by the mapping $\sigma : x \in U \to \sigma(x) \in \Pi^{-1}(U)$.

Given a connection Γ with the connection form ω in P, for every cross section $\sigma(x)$ of the bundle we can define a $\mathcal{G}$-valued connection form $A^{(\sigma)}$ *on* U by

D-3:

$$A^{(\sigma)}(\xi_\mu) \equiv A^{(\sigma)i}_{\ \mu}\, \xi_i = \omega(\sigma \cdot \xi_\mu) = \omega^i(\sigma \cdot \xi_\mu)\xi_i.$$

Notice that $A^{(\sigma)}$ *depends* upon the choise of the cross section $\sigma(x)$. $A^{(\sigma)}$ can be viewed as a connection form on the cross section $\sigma(x)$ since U is diffeomorphic to $\sigma(x)$.

Now we want to identify $A^{(\sigma)i}_{\ \mu} = \omega^i(\sigma \cdot \xi_\mu)$ as the vector potentials. To do this we will first show that they transform under different choice of cross section precisely the way vector potentials transform under gauge transformations.

First, observe that any two cross sections can be related by a transition function $a(x) \in G$, i.e.,

$$\forall\, \sigma(x),\ \sigma'(x),\ \exists\ a(x) \in G$$
$$\sigma'(x) = \sigma(x) \cdot a(x) = R_{a(x)}\sigma(x).$$

Now, for $\xi_\mu \in T_x(M)$ let $a \cdot \xi_\mu \in T_{a(x)}(G)$ be the vector induced by $a(x) : x \in U \to a(x) \in G$, where $T_x(M)$ and $T_{a(x)}(G)$ are the tangent spaces of M at x and of G at $a = a(x)$. Then $\sigma' \cdot \xi_\mu$ is the image of $(\sigma \cdot \xi_\mu, a \cdot \xi_\mu) \in T_{\sigma(x)}(P)$ $+ T_{a(x)}(G)$ under the mapping $(\sigma(x), a(x)) \in P \times G \to \sigma'(x)$ $\in P$. From Leibniz' rule, we have

$$\begin{aligned} \sigma' \cdot \xi_\mu &= (\sigma \cdot \xi_\mu) \cdot a(x) + \sigma(x) \cdot (a \cdot \xi_\mu) \\ &= R_{a(x)} \cdot (\sigma \cdot \xi_\mu) + (\sigma'(x) \cdot a^{-1}(x)) \cdot (a \cdot \xi_\mu) \\ &= R_{a(x)} \cdot (\sigma \cdot \xi_\mu) + \sigma'(x) \cdot (L_{a^{-1}(x)} a \cdot \xi_\mu), \end{aligned}$$

where σ, when acting on a vector $a \cdot \xi_\mu$ in $T(G)$, is viewed as the mapping from $T_a(G)$ to $T_{\sigma(x)\cdot a}(P)$ induced by $\sigma(x) : a \in G \to \sigma(x) \cdot a \in P$ and similarly for σ' acting on $L_{a^{-1}(x)} a \cdot \xi_\mu$.[5] So, taking the values of ω on both sides of the equality, we obtain

$$\begin{aligned} \xi_i A^{(\sigma')i}_{\ \mu} &\equiv \omega(\sigma' \cdot \xi_\mu) \\ &= \xi_i \omega^i [R_{a(x)} \cdot (\sigma \cdot \xi_\mu) + \sigma'(x) \cdot L_{a^{-1}(x)} a \cdot \xi_\mu] \\ &= [\mathrm{ad}(a^{-1}(x)) \cdot \xi_i] \omega^i(\sigma \cdot \xi_\mu) + \omega(\sigma'(x) \cdot L_{a^{-1}(x)} a \cdot \xi_\mu), \end{aligned} \tag{5}$$

where we have used the fact that ω is linear and have taken into account the property (b) of D-2′.

We now notice that ω reverses the mapping σ'. Indeed if ζ is the left invariant vector field on G which is equal to $a \cdot \xi_\mu$ at $a(x) \in G$, then the value of ζ at the origin $e \in G$ is $L_{a^{-1}(x)} a \cdot \xi_\mu$ and $\sigma' \cdot (L_{a^{-1}(x)} a \cdot \xi_\mu)$ is the value of the fundamental vector field ζ^* at $\sigma'(x) \in P$ and hence $\omega(\sigma'(L_{a^{-1}(x)} a \cdot \xi_\mu)) = L_{a^{-1}(x)} a \cdot \xi_\mu$. By using this and the definition D-3, Eq. (5) can be rewritten in the form

$$\xi_i A^{(\sigma')i}_{\ \mu} = [\mathrm{ad}(a^{-1}) \cdot \xi_i] \cdot A^{(\sigma)i}_{\ \mu} + L_{a^{-1}(x)} a \cdot \xi_\mu. \tag{5′}$$

Now, by observing that in the adjoint representation

$$\xi_j = (\xi_j)^i_{\ k} = f^i_{jk}$$

and

$$a(x) = e^{t_j \theta^j(x)},$$

where $\theta^j(x)$ $(j = 4+1 \cdots 4+n)$ are the canonical coordinates of $a(x) \in G$, Eq. (5′) can be written as

$$\xi_i A^{(\sigma')i}_{\ \mu} = e^{-t_j\theta^j(x)} \xi_i e^{t_j\theta^j(x)} A^{(\sigma)i}_{\ \mu} + e^{-t_j\theta^j(x)} \partial_\mu (e^{t_j\theta^j(x)}). \tag{5″}$$

We recognize this as the familiar transformation law of the vector potentials in gauge theory.

Conversely, by running the argument backwards one can verify that for every family of $\mathcal{G}$-valued 1-forms $\{A^{(\sigma)}\}$ on U of M satisfying the condition (5″), there exists a unique connection which gives rise to $\{A^{(\sigma)}\}$ by the prescription D-3.[7]

In short, the gauge dependent vector fields A^i_μ in gauge theory are nothing more than the coefficients of the cross section dependent connection forms on U. A choice of a cross section corresponds to a choice of gauge in this picture. This was first pointed out by Trautmen.[8]

4. CHOICE OF A METRIC

We will now introduce a metric on the bundle space so that we can establish there a Riemann geometry.

We assume that M is a metric manifold with a metric $g_{\mu\nu}$. For G we choose the invariant metric g_{ik}[9]

$$g_{ik} \equiv f^m_{in} f^n_{mk}. \tag{6}$$

We will assume G to be semisimple so that the metric defined above can be inverted. We call a metric γ_{ab} on

the bundle P compatible with the metrics $g_{\mu\nu}$ and g_{ik} on M and G if $\gamma_{ab}\hat{\xi}^a_\mu\hat{\xi}^b_\nu = g_{\mu\nu}$ and $\gamma_{ab}\xi^{*a}_i\xi^{*b}_k = g_{ik}$. If, in addition to the compatibility with the metrics $g_{\mu\nu}$ and g_{ik}, we require the metric γ_{ab} to make the horizontal and vertical subspaces orthogonal to each other, then the metric γ_{ab} is unique. This metric is given by

D-4: (1) $\gamma_{ab}\hat{\xi}^a_\mu\hat{\xi}^b_\nu = g_{\mu\nu}$,

(2) $\gamma_{ab}\xi^a_\mu\xi^{*b}_k = 0$,

(3) $\gamma_{ab}\xi^{*a}_i\xi^{*b}_k = g_{ik}$.

Notice that this definition is independent of the choice of a basis. Such a metric has been previously constructed by Kaluza[1] for the Abelian (with $g_{55}=1$) and by deWitt and Kerner[2] for the non-Abelian cases respectively.

We now consider the Riemann geometry of the bundle with the so-defined metric.

5. CHOICE OF A BASIS

We still did not prove that $F^k_{\mu\nu}$ can be viewed as Yang–Mills fields. To show this and to calculate the curvature of the bundle space, we now introduce a basis for the bundle.

Remember that from (c) of D-1, the bundle is locally trivial and we can always parametrize locally $p \in \Pi^{-1}(U)$ by $(\Pi(p), \phi(p)) \equiv (x, a) \in U\times G$.

In this parametrization $\Pi^{-1}(U)$ is diffeomorphic to the direct product space $U\times G$, and we will write $p=(x,a)$ whenever convenient.

Given any bases of U and G, one can always introduce a basis in $\Pi^{-1}(U)$, the local direct product basis of the two bases of U and G. In particular, one can take the local direct product basis $\bar{\xi}_\mu$ $(\mu=1\cdots4)$ and $\bar{\xi}_i$ $(i=5\cdots4+n)$ of the coordinate basis ξ_μ $(\mu=1\cdots4)$ of U and the left invarient basis ξ_i $(i=4+1\cdots4+n)$ of G for our basis of $\Pi^{-1}(U)$.

We clearly have $\bar{\xi}_i = \xi^*_i$ $(i=5\cdots4+n)$. But notice that $\bar{\xi}_\mu \neq \hat{\xi}_\mu$ (the horizontal lift). This is so because *for every* $a\in G$ the $\bar{\xi}_\mu$ $(\mu=1\cdots4)$ form a basis of the tangent space of the trivial cross section $\sigma_a(x)$ defined as the set of points $p=(x,a)$ *with fixed* a, which is by itself a smooth submanifold. Consequently, $\bar{\xi}_\mu$ $(\mu=1\cdots4)$ should form a closed Lie algebra within themselves. On the other hand $\hat{\xi}_\mu$ in general do not by themselves form a closed Lie algebra: The commutator $[\hat{\xi}_\mu, \hat{\xi}_\nu]$, as we have seen in Eq. (1), does acquire a vertical component.

A connection which gives vanishing vertical component to all $[\hat{\xi}_\mu, \hat{\xi}_\nu]$ is called the flat connection, in which case $F^k_{\mu\nu}$ vanish identically although the cross section dependent vector potentials $A^{(\sigma)i}_\mu$ need not vanish.

From their definition, the commutation relations of $\bar{\xi}_a$ now as vector fields in the bundle are

$$\begin{aligned}&[\bar{\xi}_i, \bar{\xi}_j] = f^k_{ij}\bar{\xi}_k,\\ &[\bar{\xi}_i, \bar{\xi}_\mu] = [\bar{\xi}_i, \partial_\mu] = 0,\\ &[\bar{\xi}_\mu, \bar{\xi}_\nu] = [\partial_\mu, \partial_\nu] = 0.\end{aligned} \tag{7}$$

Now from the definition D-3 we can attach gauge potentials $A^{(\sigma_a)i}_\mu$ to the family of trivial cross sections $\sigma_a(x)$, $a\in G$:

$$A^{(\sigma_a)i}_\mu = \omega^i(\bar{\xi}_\mu(x,a)). \tag{8}$$

These trivial cross sections ("gauges") $\sigma_a(x)$ form a family parametrized by the group element a, i.e., to every $a\in G$ there corresponds such a cross section. In other words, for every a, $\sigma_a(x)$ determines a gauge.

For simplicity we introduce the notation

$$B^i_\mu(x,a) \equiv A^{(\sigma_a)i}_\mu = \omega^i(\bar{\xi}_\mu(x,a)). \tag{8'}$$

As a consequence of Eq. (4), the horizontal lift $\hat{\xi}_\mu$ can be written in this local direct product basis as (see Fig. 2)

$$\begin{aligned}\hat{\xi}_\mu &= \bar{\xi}_\mu - \bar{\xi}_i\omega^i(\bar{\xi}_\mu)\\ &= \partial_\mu - \bar{\xi}_i B^i_\mu(x,a)\\ &\equiv D_\mu.\end{aligned} \tag{9}$$

From the group transformation properties of the connection form ω specified by property (b) of D-2′, we have

$$\bar{\partial}_i B^k_\mu(x,a) = \partial^*_i B^k_\mu = -f^k_{ij}B^j_\mu(x,a). \tag{10}$$

From Eqs. (9) and (10) we can reconfirm the commutation relations (1):

$$\begin{aligned}&[\xi^*_i, \xi^*_j] = [\bar{\xi}_i, \bar{\xi}_j] = f^k_{ij}\bar{\xi}_k = f^k_{ij}\xi^*_k,\\ &[\xi^*_i, \hat{\xi}_\mu] = [\bar{\xi}_i, \bar{\xi}_\mu - \bar{\xi}_j B^j_\mu] = [\bar{\xi}_i, D_\mu] = 0,\\ &[\hat{\xi}_\mu, \hat{\xi}_\nu] = [\bar{\xi}_\mu - \bar{\xi}_i B^i_\mu, \bar{\xi}_\nu - \bar{\xi}_j B^j_\nu] = [D_\mu, D_\nu]\\ &\qquad = -F^k_{\mu\nu}\bar{\xi}_k = -F^k_{\mu\nu}\xi^*_k,\end{aligned} \tag{1'}$$

where *now* $F^k_{\mu\nu}$ *are determined in terms of the potentials* B^k_μ *by*

$$F^k_{\mu\nu} = \partial_\mu B^k_\nu - \partial_\nu B^k_\mu + f^k_{ij}B^i_\mu B^j_\nu,$$

the usual Yang–Mills form.

One can also derive from Eq. (10) the gauge covariance of $F^k_{\mu\nu}$:

$$\partial^*_i F^k_{\mu\nu}(x,a) = \bar{\partial}_i F^k_{\mu\nu}(x,a) = -f^k_{ij}F^j_{\mu\nu}(x,a). \tag{2'}$$

Thus we have proved our claim that $F^k_{\mu\nu}$ do correspond to Yang–Mills fields.

From one of the Jacobi identies

$$[\hat{\xi}_\rho, [\hat{\xi}_\mu, \hat{\xi}_\nu]] + [\hat{\xi}_\mu, [\hat{\xi}_\nu, \hat{\xi}_\rho]] + [\hat{\xi}_\nu, [\hat{\xi}_\rho, \hat{\xi}_\mu]] = 0,$$

we have

$$D_\rho F^k_{\mu\nu} + D_\mu F^k_{\nu\rho} + D_\nu F^k_{\rho\mu} = 0 \tag{11}$$

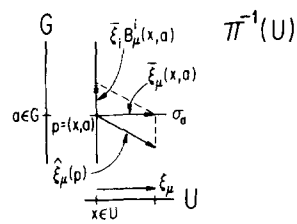


FIG. 2. Relation between the local direct product basis $\bar{\xi}_a$ and the horizonal lift basis $\hat{\xi}_a$.

for every μ, ν, ρ, and k, which are the well-known gauge Bianchi identies. In particular, they guarantee the absence of non-Abelian magnetic monopoles.

6. THE CURVATURE OF THE BUNDLE

Since the bundle space P has been organized into a $(4+n)$-dimensional Riemann space, we can calculate the curvature of the bundle P. In the local direct product basis the metric defined by D-4 can be written explicitly as

$$\gamma_{ab}=\left(\begin{array}{c|c} g_{\mu\nu}+g_{ik}B^i_\mu B^k_\nu & B^i_\mu g_{ik} \\ \hline g_{ik}B^k_\nu & g_{ik} \end{array}\right) \tag{12}$$

and

$$\gamma^{ab}=\left(\begin{array}{c|c} g^{\mu\nu} & -g^{\mu\nu}B^k_\nu \\ \hline -B^i_\mu g^{\mu\nu} & g^{ik}+g^{\mu\nu}B^i_\mu B^k_\nu \end{array}\right),$$

where

$$g^{\mu\nu}g_{\nu\rho}=\delta^\mu{}_\rho, \quad g^{ij}g_{jk}=\delta^i{}_k$$

and

$$\gamma^{ab}\gamma_{bc}=\delta^a{}_c.$$

To calculate the curvature, one has to keep in mind that we are using a noncoordinate basis here and must take into account the nonvanishing commutators.

In the torsion free theory the formulas for the Christoffel coefficients Γ^a_{bc} and the curvature tensor $R_{abc}{}^d$ in a noncoordinate basis $\hat{e}_a$ $(a=1\cdots 4+n)$ with the metric γ_{ab} are given by[10]

$$\begin{aligned}\Gamma^a_{bc}&=\tfrac{1}{2}\gamma^{ad}(\partial_b\gamma_{cd}+\partial_c\gamma_{bd}-\partial_d\gamma_{bc}-c^e_{bd}\gamma_{ce}-c^e_{cd}\gamma_{be})\\&\quad+\tfrac{1}{2}c^a_{bc},\\ R_{abc}{}^d&=c^e_{ab}\Gamma^d_{ec}-\partial_a\Gamma^d_{bc}+\partial_b\Gamma^d_{ac}-\Gamma^d_{ae}\Gamma^e_{bc}+\Gamma^d_{be}\Gamma^e_{ac},\end{aligned} \tag{13}$$

where ∂_a is meant as the directional derivative $\partial_{\hat{e}_a}$ and the c^a_{bc} are the coefficients of the commutation relations of the (noncoordinate) basis $\hat{e}_a$:

$$[\hat{e}_a,\hat{e}_b]=c^c_{ab}\hat{e}_c.$$

We assume, for simplicity, the torsion-free Christoffel coefficients for the bundle. In the local direct products basis $\hat{e}_a=\tilde{\xi}_a$ and from Eq. (7) we determine the commutation coefficients:

$$\begin{aligned}&c^k_{ij}=f^k_{ij}\\&c^\mu_{ij}=c^k_{i\nu}=c^\mu_{i\nu}=c^k_{\mu\nu}=c^\alpha_{\mu\nu}=0.\end{aligned} \tag{14}$$

Using Eqs. (12), (13) and (14), we find[11]

$$\begin{aligned}&\bar{\Gamma}^i_{jk}=\tfrac{1}{2}f^i_{jk}=\Gamma^i_{jk},\\&\bar{\Gamma}^\alpha_{jk}=0,\\&\bar{\Gamma}^\alpha_{k\mu}=\bar{\Gamma}^i_{k\mu}=\tfrac{1}{2}(f^i_{jk}B^j_\mu+g_{kl}g^{\alpha\beta}B^i_\alpha F^l_{\beta\mu}),\\&\bar{\Gamma}^\alpha_{\mu k}=\Gamma^\alpha_{k\mu}=\tfrac{1}{2}g^{\alpha\beta}g_{kl}F^l_{\mu\beta},\\&\bar{\Gamma}^i_{\mu\nu}=\tfrac{1}{2}(\partial_\mu B^i_\nu+\partial_\nu B^i_\mu)+\tfrac{1}{2}g^{\alpha\beta}g_{ik}B^i_\alpha(F^j_{\beta\mu}B^k_\nu+F^j_{\beta\nu}B^k_\mu)\\&\qquad-\Gamma^\alpha_{\mu\nu}B^i_\alpha,\\&\bar{\Gamma}^\alpha_{\mu\nu}=\Gamma^\alpha_{\mu\nu}+\tfrac{1}{2}g^{\alpha\beta}g_{ik}(B^i_\mu F^k_{\nu\beta}+B^i_\nu F^k_{\mu\beta}),\end{aligned} \tag{15}$$

where Γ^i_{jk} and $\Gamma^\alpha_{\mu\nu}$ are the torsion free Christoffel coefficients of G and M with the metrics g_{ik} and $g_{\mu\nu}$, and

$$\begin{aligned}\bar{R}_{ik}&=R_{ik}+\tfrac{1}{4}g_{ij}g_{kl}g^{\mu\alpha}g^{\nu\beta}F^j_{\mu\nu}F^l_{\alpha\beta},\\ \bar{R}_{\mu k}&=\bar{R}_{k\mu}=\tfrac{1}{4}g_{kl}B^l_\mu+\tfrac{1}{4}g_{kl}g_{ij}g^{\alpha\gamma}g^{\beta\delta}B^i_\mu F^j_{\alpha\beta}F^l_{\gamma\delta}\\&\quad+\tfrac{1}{2}g_{kl}g^{\alpha\beta}\nabla_\alpha F^l_{\mu\beta},\\ \bar{R}_{\mu\nu}&=R_{\mu\nu}+\tfrac{1}{4}g_{ik}B^i_\mu B^k_\nu+\tfrac{1}{4}g_{ij}g_{kl}g^{\alpha\gamma}g^{\beta\delta}B^i_\mu B^k_\nu F^j_{\alpha\beta}F^l_{\gamma\delta}\\&\quad-\tfrac{1}{2}g^{\alpha\beta}g_{ik}F^i_{\mu\alpha}F^k_{\nu\beta}+\tfrac{1}{2}g_{ik}g^{\alpha\beta}(B^i_\mu\nabla_\alpha F^k_{\nu\beta}+B^i_\nu\nabla_\alpha F^k_{\mu\beta}),\end{aligned} \tag{16}$$

where $R_{ik}=\frac{1}{4}g_{ik}$ and $R_{\mu\nu}$ are the Ricci tensors of G and M, and ∇_α is the gauge *and* generally covariant derivative, e.g.,

$$\nabla_\alpha F^k_{\mu\nu}=\partial_\alpha F^k_{\mu\nu}-\Gamma^\beta_{\alpha\mu}F^k_{\beta\nu}-\Gamma^\beta_{\alpha\nu}F^k_{\mu\beta}+f^k_{ij}B^i_\alpha F^j_{\mu\nu}.$$

Finally we find

$$\begin{aligned}\bar{R}&=g^{\mu\nu}R_{\mu\nu}+g^{ik}R_{ik}-\tfrac{1}{4}g_{ik}g^{\mu\alpha}g^{\nu\beta}F^i_{\mu\nu}F^k_{\alpha\beta}\\&\equiv R_M+R_G-\tfrac{1}{4}F^2.\end{aligned} \tag{17}$$

7. UNIFIED ACTION INTEGRAL

The Einstein—Hilbert action integral I_{n+4} of the bundle space P is written as

$$I_{n+4}=\int\sqrt{-\gamma}\;\bar{R}\,d^4x\,d^nG. \tag{18}$$

It can be formally integrated out over the group degree of freedom to give a four-dimensional action integral I_4:

$$I_4=V_G\times\int\sqrt{-g_M}(R_M+R_G-\tfrac{1}{4}F^2)d^4x \tag{19}$$

(V_G = the volume of the group manifold = const), which is postulated as the unified action.

Notice that if the space—time manifold M is not flat, the constant R_G plays the nontrivial role of a cosmological constant in the theory. This will be discussed in more detail in Sec. 9.

8. OTHER BASES

We have proved our statement by choosing an explicit basis, the local direct product basis of the bundle defined by Eqs. (7), which corresponds to choosing a specific class of gauges. Of course, one can choose different bases and obtain the same result.

For example, given any arbitrary cross section $\sigma(x)$ on U, one can span the whole $\Pi^{-1}(U)$ by the family of cross sections $R_a\cdot\sigma$ with every $a\in G$. Now one can choose a basis for $\Pi^{-1}(U)$ in the following way. Choose $\tilde{\xi}_i\equiv\xi^*_i(i=5\cdots 4+n)$ and define $\tilde{\xi}_\mu$ $(\mu=1\cdots 4)$ at each point of $p\in\Pi^{-1}(U)$ to be the mapping of $\xi_\mu=\partial_\mu$ $(\mu=1\cdots 4)$ into the tangent space T_p of P by one of $R_a\cdot\sigma: x\in U\to p\in\Pi^{-1}(U)$.

Clearly these $\tilde{\xi}_a$ form a basis for the tangent spaces of $\Pi^{-1}(U)$ and their commutation relations are

$$\begin{aligned}&[\tilde{\xi}_i,\tilde{\xi}_j]=f^k_{ij}\tilde{\xi}_k,\\&[\tilde{\xi}_i,\tilde{\xi}_\mu]=0,\\&[\tilde{\xi}_\mu,\tilde{\xi}_\nu]=0.\end{aligned} \tag{20}$$

It is straightforward to retrace every previous step in this basis and to obtain the same result. We will not duplicate the arguments here.

It is worthwhile mentioning that there exists yet another basis where the calculations become particularly simple. Notice that $\hat{\xi}_\mu$ $(\mu=1\cdots4)$ and $\hat{\xi}_i \equiv \xi_i^*$ $(i=5\cdots4+n)$ can also be used as a basis for the bundle with the commutation relations (1). We will call this basis the horizontal lift basis.

In this horizontal lift basis γ_{ab} can be written as

$$\gamma_{ab}=\left(\begin{array}{c|c} g_{\mu\nu} & 0 \\ \hline 0 & g_{ik} \end{array}\right) \tag{21}$$

and

$$\gamma_{ab}=\left(\begin{array}{c|c} g^{\mu\nu} & 0 \\ \hline 0 & g^{ik} \end{array}\right).$$

Again assuming the torsion-free Christoffel coefficients for simplicity, we find in this horizontal lift basis

$$\begin{aligned}
&\hat{\Gamma}^i_{jk}=\tfrac{1}{2}f_{jk}^i=\Gamma^i_{jk},\\
&\hat{\Gamma}^\alpha_{jk}=0,\\
&\hat{\Gamma}^i_{\mu k}=\hat{\Gamma}^i_{k\mu}=0,\\
&\hat{\Gamma}^\alpha_{\mu k}=\hat{\Gamma}^\alpha_{k\mu}=\tfrac{1}{2}g^{\alpha\beta}g_{kl}F^l_{\mu\beta},\\
&\hat{\Gamma}^i_{\mu\nu}=-\tfrac{1}{2}F^i_{\mu\nu},\\
&\hat{\Gamma}^\alpha_{\mu\nu}=\Gamma^\alpha_{\mu\nu}=\tfrac{1}{2}g^{\alpha\beta}(\partial_\mu g_{\nu\beta}+\partial_\nu g_{\mu\beta}-\partial_\beta g_{\mu\nu}),
\end{aligned} \tag{22}$$

and

$$\begin{aligned}
&\hat{R}_{ik}=R_{ik}+\tfrac{1}{4}g_{ij}g_{kl}g^{\alpha\beta}g^{\gamma\delta}F^j_{\alpha\gamma}F^l_{\beta\delta},\\
&\hat{R}_{\mu k}=\hat{R}_{k\mu}=\tfrac{1}{2}g_{kl}g^{\alpha\beta}\nabla_\alpha F^l_{\mu\beta},\\
&\hat{R}_{\mu\nu}=R_{\mu\nu}-\tfrac{1}{2}g_{ik}g^{\alpha\beta}F^i_{\mu\alpha}F^k_{\nu\beta},
\end{aligned} \tag{23}$$

and

$$\hat{R}=\bar{R}=R_M+R_G-\tfrac{1}{4}F^2. \tag{24}$$

Again we get the same result we had before. Notice that, although the components of tensor quantities do depend upon the choice of a basis, the curvature R being a scalar does not depend upon the choice of a basis.

In this horizontal lift basis $\hat{\xi}_\mu$ are not the basis vectors of the tangent space to any cross section $\sigma(x)$, because they do not commute under commutation. Potentials $A_\mu^{(\sigma)i}=\omega^i(\sigma\cdot\xi_\mu)$ can be introduced as before, since they are defined independent of choice of a basis in the bundle space. The horizontal lift basis is very convenient for calculations.

9. DISCUSSION

We thus have combined gravitation with gauge theory in the context of a unified geometric theory in the bundle space with the metric D-4.

In the process of the argumant we have assumed the torsion-free Christoffel connection for our bundle P and the space–time independent metric $\gamma_{ab}\xi_i^{*a}\xi_k^{*b}=g_{ik}$ for the fiber space. In general one can relax these assumptions and obtain different variants of the theory.

We now want to comment on the cosmological term[12] $R_G\sqrt{-g_M}$ in the action Eq. (19). The only dimensional constants in the theory are the universal gravitational constant G and the typical length in group space (i.e., essentially the magnitude of g_{ij}). Using units $\hbar=c=1$, one can therefore consider the Planck length $L_p=\sqrt{G}$ and the length in group space L_G. If L_G is left arbitrary, then so is the cosmological term. It is not clear what dynamical considerations are to determine it. If L_G were to be of the order of the Planck length, so that G be the only dimensional parameter, one would obtain a by far $(\sim10^{120})$ too large cosmological term. In that case, one might try a metric for G which gives a vanishing R_G and remove the cosmological term altogether. Otherwise, one can add a countercosmological term "by hand."

We also emphasize at this point that the universal coupling constant g of the Yang–Mills field is also arbitrary in this theory. For simplicity we have set $g=1$ in this paper. The theory for arbitrary g is then obtained by systematically replacing $A_\mu^{(\sigma)i}$ and B_μ^i by $gA_\mu^{(\sigma)i}$ and gB_μ^i.

One might also try the Jordan–Thirry version[13] of the theory by letting $\gamma_{ab}\xi_i^{*a}\xi_k^{*b}=\phi_{ik}(x)$ become scalar fields.[14]

Beyond these technicalities, however, we would like to emphasize more the general structure of the theory: the combined geometry of the space–time with the internal symmetry space. In this geometry, once an internal symmetry is given, one is led to have gauge fields without referring to the existence of any "matter" fields.

Of course, one can always introduce matter fields in our geometry by hand by introducing another bundle space: the fiber bundle associated to the principal fiber bundle.[15] It would be very interesting if one could obtain the Fermi fields as a part of the gauge fields of some supersymmetry.[16]

It is well known that the usual theory of gravitation itself can be viewed as a theory of connections of the bundle P with the space-time as the base manifold and $G=O(1,3)$ as the structural group.[17] In this picture the Christoffel coefficients $\Gamma^\rho_{\mu\nu}$ appear as the cross section dependent connection forms on U with $G=O(1,3)$. This explains the well-known nontensorial transformation properties of $\Gamma^\rho_{\mu\nu}$. One is led to Einstein's theory from the torsion-free connection, Cartan's theory[18] from the torsional connection.[19] Utiyama's and Kibble's work[20] can be argued more elegantly and consistently in this picture.

ACKNOWLEDGMENTS

It is a great pleasure to thank Professor P.G.O. Freund for suggesting this problem and for the continuous advice which was essential in completing this paper. I am grateful to Professor Y. Nambu for careful reading of the manuscript and Professor R. Lashof in the Department of Mathematics of the University of Chicago for helpful discussions. I am also indebted to R.M. Wald, J. Friedman, and P.S. Jang for many discussions. Finally, I would like to thank S.O. Ahn for encouragement.

*Work supported in part by the National Science Foundation, Contract No. MPS75-08833.

[1]T. Kaluza, Sitzber. Preuss. Akad. Wiss., 966 (1921).
[2]O. Klein, Z.Physik **37**, 895 (1926).
[3]B. DeWitt, *Dynamical Theories of Groups and Fields* (Gordon and Breach, New York, London, Paris, 1965), p. 139; R. Kerner, Ann. Inst. H. Poincaré **9**, 143 (1968); A. Trautman, Rep. Math. Phys. **1**, 29 (1970).
[4]For general mathematical background we refer to S. Sternberg, *Lectures on Differential Geometry* (Prentice-Hall, Englewood Cliff, N.J., 1964); S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963); K. Nomizu, *Lie Groups and Differential Geometry*, Publ. Math. Soc. Japan #2, (Math. Soc. Japan, Tokyo, 1956); R. L. Bishop and R. J. Crittenden, *Geometry of Manifolds* (Academic, New York and London, 1965).
[5]For simplicity we will use in this paper the same notation for a mapping of a manifold to another and for the corresponding mapping of a vector (i.e., an element of the tangent space) to another induced by the mapping of the manifolds. For example, if L_a hits an element of the group G, it should be interpreted as a mapping of G into itself, and if L_a hits a vector, it should be interpreted as a mapping of the vector to another vector induced by the mapping L_a of G.
[6]For a noncoordinate basis, see for example, C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[7]See, for example, S. Kobayashi and K. Nomizu, Ref. 4, p. 66.
[8]A. Trautman, Ref. 3.
[9]Of course, $g_{\mu\nu}$ and g_{ik} are in terms of the previously chosen coordinate basis of M and the left invariant basis of G.
[10]See, for example, C.W. Misner, K.S. Thorne, and J.A. Wheeler, Ref. 6.
[11]Our argument in this section is similar to that of R. Kerner, Ref. 3. However, our results [Eq. (14), (15), and (16)] differ from his.
[12]As for cosmological models with a cosmological constant, see, for example, H. Bondi, *Cosmology* (Cambridge U.P., Cambridge, 1960), 2nd ed., Chap. IX. See also S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972), pp. 155, 613–16.
[13]P. Jordan, Z. Phys. **157**, 112 (1959); Y.R. Thirry, C. R. Acad. Sci. Paris **226**, 216 (1948); see also, C. Brans and R.H. Dicke, Phys. Rev. **124**, 925 (1961).
[14]Y.M. Cho and P.G.O. Freund, to be published.
[15]See, for example, A. Trautman, Ref. 3; K. Nomizu, Ref. 4, or S. Kobayashi and K. Nomizu, Ref. 4.
[16]For a similar approach but in somewhat different context, see D.V. Volkov and V.P. Akulov, Zh. Eksp. Teor. Fiz. Pis'ma Red. **16**, 621 (1972) [JETP Lett. **16**, 438 (1972)]; D.V. Volkov and V.A. Soroka, *ibid.* **18**, 529 (1973) [*ibid.* **18**, 312 (1973)].
[17]For a precise mathematical justification, see S. Kobayashi and K. Nomizu, Ref. 4, p. 158.
[18]E. Cartan, Ann. École Normale **40**, 325 (1923); **41**, 1 (1924).
[19]T.W. B. Kibble, J. Math. Phys. **2**, 212 (1960).
[20]R. Utiyama, Phys. Rev. **101**, 1592 (1956); T.W.B. Kibble, Ref. 19.

Exact quantization of the nonlinear Schrödinger equation*

D. J. Kaup

Department of Physics, Clarkson College of Technology, Potsdam, New York 13676
(Received 10 March 1975)

By means of an "inverse scattering transform," we can exactly quantize the one-dimensional nonlinear Schrödinger equation $i\Psi_t = -(\hbar^2/2m)\Psi_{xx} - \epsilon^2(\Psi^*\Psi)\Psi$ for any value of ϵ^2 = real. When $\epsilon^2 < 0$, the eigenvalues of the number operator, field momentum operator, and the Hamiltonian are found to be *exactly* the same as the linear case. In other words, by quantizing the exact theory, no effects corresponding to "renormalization" are found, and the zero point energy is independent of ϵ^2. When $\epsilon^2 < 0$, the Hamiltonian is unbounded from below, and, in addition to the above spectra of eigenvalues, bound states can occur. Each bound state can be interpreted to be a bound state of n "excitations," moving in a coherent fashion and with a binding energy proportional to the cube of the number of excitations. This problem is also formally equivalent to the N-body problem with a delta-function interaction solved by Bethe, with which we shall contrast our results, and we shall conclude by making certain remarks concerning ordinary field quantization versus "scattering space" quantization.

1. INTRODUCTION

It is generally recognized that if an equation can be solved exactly in the classical limit (with the possible exception of the anharmonic oscillator), then it can also be quantized exactly. With the advent of a very powerful transform, the "inverse scattering transform",[1] which was pioneered by Gardner, Green, Kruskal, and Miura,[2] a wide class of one-dimensional nonlinear dispersive[1-12] (and nondispersive[13]) evolution equations can now be solved exactly in the classical domain. Among these nonlinear equations, there are two which can serve as one-dimension models for quantized field theories. These are the "nonlinear Schrödinger equation,"[3] which is nonrelativistic, and the sine–Gordon equation,[5,11,12] which is relativistic. Of these two, it is the nonlinear Schrödinger which is a central equation, since it is the nonrelativistic and weak coupling limit of not only the nonlinear Klein–Gordon equation,[14] the sine–Gordon equation, but also, of any general one-component field theory whose lowest order nonlinearity is a cubic term. In this respect, the nonlinear Schrödinger equation occupies a central position, since the quantization of this field will be a limit for many other theories.

Although much work has gone into an analysis of these model field theories and their perturbation expansions, up until now, there was never any serious hope of ever being able to solve any of these models exactly. However, with the advent of the "inverse scattering transform," not only are we able to obtain exact solutions in the classical limit, but it is also possible now to quantize the same systems exactly. The particular reason why the ability to quantize immediately follows in general upon obtaining a classical solution, is related to the nature of the inverse scattering transform. By using this transform, one can map (nonlinearily!) a field into a new space, and in this new space, the evolution in time of the field is described by *linear*, separable equations. In other words, under this mapping, the field has been decomposed into its "normal modes," and each normal mode evolves in time independently of all other normal modes. Thus, one can immediately identify the action-angle variables, obtain the Hamiltonian in terms of these action variables, and, except for possible complications due to the phase-space topology, one can immediately apply the canonical quantization procedure and obtain a quantum theory. Note the analogy of this method with the method of quantizing *linear* field theories. For linear fields, one uses the Fourier transform to map (this time linearly) the field into a new space ("Fourier space") where the evolution in time of the field is also described by linear separable equations. Under the Fourier transformation, the linear field has also been mapped into its normal modes, and it is these modes which are then quantized. And, in several other ways, the analogy between this nonlinear inverse scattering transform and the linear Fourier transform is so close, that one can also consider the inverse scattering transform to be simply a "nonlinear Fourier transform."[1]

Due to the strong analogy pointed out above, for those equations which can be solved exactly in the classical limit by an inverse scattering transform, we propose to quantize directly the normal modes, and *not* the fields. We do this for two reasons. First, quantization of the normal modes is extremely simple since they are all linearly independent, and many times, it is the simplest method which has the most utilitary value. Second, we appeal to the Bohr correspondence principle. In brief, if, in the limit of $\hbar \to 0$, the evolution of the appropriate matrix elements obeys the classical equations of motion, then we have obtained *a* quantum theory. Thus, if this theory does differ from that obtained by quantizing the fields, and if both theories satisfy the correspondence principle, then they can only differ by terms of higher order in $\hbar$ (i.e., can differ only by their factor ordering and thus effects due to normal ordering).

In Sec. 2, we shall briefly review the method for solving the nonlinear Schrödinger equation in the classical limit by the inverse scattering transform. Here, we shall point out the normal modes for this field, which consist of "solitons"[15] and "radiation," as well as their equations of motion. Then, upon defining the usual number density, field momentum, and field energy, we shall show how one determines these quantities in terms of the normal modes.

In Sec. 3, we shall proceed to quantize these normal modes by the usual canonical quantization procedure. However, due to the existence of certain nonholonomic constraints, care has to be used, and one must carry out certain classical canonical transformations on the action-angle variables in order to eliminate these nonholonomic constraints. But, once they are eliminated, one finds that the resulting normal modes are simply those of only either free particles or harmonic oscillators and, thus, quantization becomes trivial. Upon quantization, the soliton can be interpreted as a bound state of n "excitations," whose binding energy is proportional to $(n+\frac{1}{2})^3$. Of course, these bound states have no analogy in the second-quantized version of the *linear* Schrödinger equation. However, the "radiation" part of the spectrum does have an analogy which is one-to-one. In fact, and rather surprisingly, the eigenvalues for the number operator, momentum operator, and the Hamiltonian for the "radiation" part of the spectrum is *exactly* the same as for the second-quantized *linear* Schrödinger equation! Thus, the only difference in their eigenvalue spectra lies in the addition of the bound states for the nonlinear case, when $\epsilon^2>0$.

Finally we conclude with a comparison of our results and those for the N-body problem with a δ-function interaction.[16,17] A major point here is that, in quantizing first in scattering space, the commutation relations for the fields are not uniquely specified until we have specified the factor ordering in the direct scattering transform and the inverse scattering transform. The manner in which it could affect our results will be pointed out.

2. THE NONLINEAR SCHRÖDINGER EQUATION

First, we shall briefly present the method for solving the one-dimensional nonlinear Schrödinger equation[3] in the classical limit, and point out how this method effectively maps the field into a "nonlinear Fourier transform space," which is the "scattering data" of the Zakharov–Shabat eigenvalue problem. The scattering data will be defined, its time evolution will be discussed, as well as how each part of the scattering data affects the evolution of the field in time. Finally, we shall define the field number, field momentum, and field energy, and show how one obtains these quantities in terms of the scattering data.

Following Zakharov–Shabat,[3] if we consider the eigenvalue problem

$$v_{1x}+i\zeta v_1=(\epsilon/\hbar)m^{1/2}\Psi v_2, \tag{2.1a}$$

$$v_{2x}-i\zeta v_2=-(\epsilon/\hbar)m^{1/2}\Psi^* v_1, \tag{2.1b}$$

where ζ is the eigenvalue, v_1 and v_2 are the eigenfunctions, and let the time evolution of $v=\begin{bmatrix}v_1\\v_2\end{bmatrix}$ be given by

$$iv_{1t}=[(\hbar/m)\zeta^2-(\epsilon^2/2\hbar)\Psi^*\Psi]v_1+\epsilon m^{-1/2}(i\zeta\Psi-\tfrac{1}{2}\Psi_x)v_2, \tag{2.2a}$$

$$iv_{2t}=-\epsilon m^{-1/2}(i\zeta\Psi^*+\tfrac{1}{2}\Psi_x^*)v_1 -[(\hbar/m)\zeta^2-(\epsilon^2/2\hbar)\Psi^*\Psi]v_2, \tag{2.2b}$$

then the integrability condition for (2.1, 2) is

$$i\hbar\Psi_t=-(\hbar^2/2m)\Psi_{xx}-\epsilon^2(\Psi^*\Psi)\Psi, \tag{2.3}$$

where ϵ^2 is the coupling constant, assumed to be real, and subscripts x and t designate partial differentiation.

The classical solution of the initial value problem is given as follows.[1,3] First, the scattering data S_+ is determined at $t=0$, where

$$S_+=\{[\zeta_j,\ \rho_j]_{j=1}^J;\ \rho(\xi)\ (\xi=\text{real})\}. \tag{2.4}$$

Let Ψ satisfy[1]

$$\int_{-\infty}^{\infty}|\Psi(x,\ t)|dx<\infty, \tag{2.5}$$

and define $\phi=\begin{bmatrix}\phi_1\\\phi_2\end{bmatrix}$ to be the solution of (2.1) which satisfies the boundary condition

$$\phi\to\begin{bmatrix}1\\0\end{bmatrix}\exp(-i\zeta x)\ \text{ as } x\to-\infty, \tag{2.6a}$$

for $\mathrm{Im}(\zeta)\geq 0$. Then, as $x\to+\infty$,

$$\phi\to\begin{bmatrix}a(\zeta)\exp(-i\zeta x)\\ b(\zeta)\exp(i\zeta x)\end{bmatrix}, \tag{2.6b}$$

which defines a and b, and which also satisfies

$$\bar{a}(\zeta)a(\zeta)+\bar{b}(\zeta)b(\zeta)=1, \tag{2.7}$$

where

$$\bar{a}(\zeta)\equiv[a(\zeta^*)]^*, \tag{2.8a}$$

$$\epsilon\bar{b}(\zeta)\equiv[\epsilon b(\zeta^*)]^*. \tag{2.8b}$$

Of course, (2.7), (2.8) are in general valid off the real ζ axis only when the $\Psi(x)$ in (2.1) is on compact support.[1] The continuous spectrum ("radiation") of the scattering data is given by

$$\rho(\xi)=b(\xi)/a(\xi)\ \ (\xi=\text{real}). \tag{2.9}$$

When $\epsilon^2<0$, ϵ is then pure imaginary, (2.1) is then self-adjoint when (2.5) is satisfied and has no bound states. Thus (2.9) is the only component of the scattering data S_+. But when $\epsilon^2>0$, (2.1) can have bound states when (2.5) is sufficiently large. We define this part of the scattering data as follows. When Ψ is on compact support,[1] $\rho(\xi)$ can be extended into the entire complex ζ plane. Then the poles of $\rho(\zeta)$ [zeros of $a(\zeta)$] in the upper half ζ plane correspond to the bound states of (2.1). Let these zeros of $a(\zeta)$ be designated by $[\zeta_j]_{j=1}^J$, where J is the total number of bound states, assumed to be finite. The residue of $\rho(\zeta)$ at $\zeta=\zeta_j$ is designated by ρ_j, which completes the definition of this part of the scattering data.

Reflecting on what has happened so far, we see that, by (2.1), we have mapped the field $\Psi(x,\ t)$ into the scattering data $S_+(\zeta,\ t)$. This mapping is a nonlinear mapping, but when the field is sufficiently small (or ϵ is sufficiently small), the mapping does become essentially linear. From (2.1), (2.6), (2.9) one can show that

$$\rho(\xi,\ t)=-\epsilon m^{1/2}\hbar^{-1}\int_{-\infty}^{\infty}\Psi^*(x,\ t)\exp(-2i\xi x)dx+O(\epsilon^3), \tag{2.10}$$

which, to lowest order in ϵ, is simply the linear Fourier transform of Ψ^*. Of course, when the higher order terms are included, this mapping becomes fully nonlinear, but it still can be considered as a "nonlinear Fourier transform."

This analogy with the linear Fourier transform is even more stronger when one considers the equations of motion of the scattering data. From (2.2), (2.5), (2.6), (2.9), one can show[1,3] that the scattering data evolves in time according to

$$i\hbar(\xi_j)_t = 0 \qquad (j=1,2,\ldots,J), \tag{2.11a}$$

$$i\hbar(\rho_j)_t = -(\hbar^2/2m)(2\xi_j)^2\rho_j \tag{2.11b}$$

$$i\hbar a(\zeta)_t = 0 \quad (\mathrm{Im}\zeta \geq 0), \tag{2.11c}$$

$$i\hbar\rho(\xi)_t = -(\hbar^2/2m)(2\xi)^2\rho(\xi) \quad (\xi = \text{real}). \tag{2.11d}$$

Not only are these equations of motion separable linear equations, but note the time evolution of $\rho(\xi)$ and ρ_j. Directly from (2.3), (2.10), we would certainly expect the equation of motion for $\rho(\xi)$ to be as given by (2.11d), when one takes the limit of $\epsilon \to 0$, but (2.11d) shows that the higher order terms in (2.10) are such that (2.11d) is satisfied for all ϵ! Thus the equation of motion for $\rho(\xi)$, in the fully nonlinear case, is exactly the same as one would have in the linear case ($\epsilon = 0$) for the Fourier transform of Ψ^*, and thus $\rho(\xi)$ always evolves in time as a Fourier component.

To reconstruct the "potential" $\Psi(x, t)$ at any later time, one proceeds as follows. First, construct the function[1,3]

$$F(z;t) = (1/2\pi)\int_{-\infty}^{\infty} d\xi\,\rho(\xi;0)\exp(i\xi z + 2i\hbar m^{-1}\xi^2 t) - i\sum_{j=1}^{J}\rho_j(0)\exp(i\xi_j z + 2i\hbar m^{-1}\xi_j^2 t). \tag{2.12a}$$

Next, solve the linear integral equation

$$K(x,y;t) - F^*(x+y;t) + \int_x^{\infty} du\,K(x,u;t)\int_x^{\infty} dw\,F(u+w;t)\,F^*(w+y;t) = 0, \tag{2.12b}$$

for $K(x, y; t)$. Then $\Psi(x, t)$ will be given by

$$\Psi(x,t) = -(2\hbar/\epsilon m^{1/2})K(x,x;t). \tag{2.13}$$

Note that, in this entire procedure, only linear techniques are required.

Before continuing, let us briefly review the nature of the solution and how each piece of the scattering data affects it. First, we shall consider the bound state part of the scattering data $[\xi_j, \rho_j]_{j=1}^{J}$, and assume the continuous part to be absent [$\rho(\xi) = 0$, $\xi = \text{real}$]. In this case, each bound state corresponds to exactly one permanent traveling wave-form, called a "soliton."[15] In general, J-solitons are present, and closed form solutions are possible.[3] For the simplest case of a one-soliton solution, we have

$$\rho(\xi) = 0 \quad (\xi = \text{real}), \tag{2.14a}$$

$$J = 1, \tag{2.14b}$$

$$\zeta_1 = \xi_1 + i\eta_1 \quad (\eta_1 > 0), \tag{2.14c}$$

$$\rho_1(0) = -2i\eta_1\exp(-2\eta_1 x_0 - 2i\xi_1\bar{x}_0), \tag{2.14d}$$

where x_0 and $\bar{x}_0$ are arbitrary, real constants. Then from (2.12), (2.13),

$$\Psi(x,t) = \frac{2\hbar\eta_1}{\epsilon m^{1/2}}\cdot\frac{\exp[-2i\xi_1(x-\bar{x}_0) - 2i\hbar m^{-1}t(\xi_1^2-\eta_1^2)]}{\cosh\{2\eta_1[x - x_0 + (\hbar/m)t2\xi_1]\}}. \tag{2.15}$$

Considering (2.15), we see that the imaginary part of the eigenvalue η_1 determines the height and width of the soliton, the real part determines the velocity, and $\rho_1(0)$ determines the initial position and phase. In general, the same correspondences still hold for a J-soliton solution, with (ζ_1, ρ_1) determining the above properties of the first soliton, etc. Note the singular dependence of Ψ on the coupling constant ϵ^2. Expansion in a power series of ϵ^2 would never allow one to find this or any other soliton state.

When the continuous spectrum is present, closed form solutions are not possible, but when no solitons are present ($J = 0$), the Neumann series solution of (2.12) is absolutely convergent if $\epsilon^2 < 0$ and is also absolutely convergent when $\epsilon^2 > 0$ if $|\rho(\xi)| < 1$.[1] To study the nature of this type of a solution, we take ρ to be very small, and consider only the lowest order term. Then to lowest order

$$\Psi(x,t) \simeq (1/2\pi)\int_{-\infty}^{\infty} 2d\xi[(-\hbar/\epsilon m^{1/2})\rho^*(\xi;0)] \times\exp(-2i\xi x - 2i\hbar m^{-1}\xi^2 t), \tag{2.16}$$

where the quantity inside the bracket is essentially the Fourier transform of $\Psi(x, 0)$. Considering (2.16), we see that the solution for the continuous part of the spectrum, in the limit of $\epsilon \to 0$, develops in time like the solution for the linear problem ($\epsilon = 0$), in that it will slowly disperse and decay away. For the fully nonlinear case, we can still expect the same to occur, except for possible nonlinear decaying oscillations.[18] Due to its property of usually occurring as decaying oscillations traveling away from a disturbance, this part of the spectrum (the continuous part) is often referred to as "radiation." When both parts of the spectrum are present, solitons and radiation, then, for large times, the solution is expected to asymptotically approach a state of separated permanent traveling waveforms (solitons) traveling in a sea of decaying radiation.

As one can see by comparing (2.3) with (2.10), mapping Ψ into the scattering data S_+ has transformed the original nonlinear equation into a set of linear equations whose solution is trivial. One can show that this mapping (2.1) is in fact a *canonical* transformation[19-21] just as are the similar maps for the KdV equation[22] and the sine–Gordon equation.[12] Furthermore, it is clear from (2.10) that this map does allow the Hamiltonian–Jacobi functional differential equation to be completely separated, and thereby allows one to determine the action-angle variables.

Although there are several ways one may proceed to determine the Hamiltonian in terms of the scattering data,[12,19-22] one of the more direct ways is to utilize the fact that all of these systems has an infinity of polynomial conserved quantities.[1,3,20] This infinity of conserved quantities follows from (2.11c) and the fact that $a(\zeta)$ is an analytic function of ζ in the upper half ζ plane. Thus, if we expand $a(\zeta)$ in an asymptotic series of ζ^{-1} about $|\zeta| = \infty$ in the upper half plane, every coefficient

must give us a conserved quantity. Now, from the relation[1]

$$\ln a(\zeta)=\sum_{j=1}^{J}\ln\left(\frac{\zeta-\zeta_j}{\zeta-\zeta_j^*}\right)-\frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{d\xi}{\xi-\zeta}\ln[1+\bar\rho(\xi)\rho(\xi)], \tag{2.17a}$$

for $\mathrm{Im}(\zeta)>0$, where the sum is absent if $\epsilon^2<0$ (since $J=0$), and $\bar\rho$ is defined by

$$\epsilon\bar\rho(\xi)=[\epsilon\rho(\xi)]^*, \tag{2.17b}$$

we can obtain this expansion in terms of the scattering data S_+.

To obtain it in terms of the field, we return to (2.1) and (2.6). Then

$$\ln a(\zeta)=\lim_{x\to\infty}\ln w(\zeta,x), \tag{2.18a}$$

where

$$w(\zeta,x)=\phi_1(\zeta,x)\exp(i\zeta x), \tag{2.18b}$$

and w is given as the solution of

$$w(\zeta,x)=1+\int_{-\infty}^{x}M(\zeta,x,y)w(\zeta,y)\,dy, \tag{2.18c}$$

and

$$M(\zeta,x,y)=-(\epsilon^2m/\hbar^2)\Psi^*(y)\int_y^x\exp[2i\zeta(z-y)]\Psi(z)\,dz. \tag{2.18d}$$

The Neumann series solution of (2.18c) is absolutely convergent in the upper half ζ plane.[1,3] Expanding (2.17) and (2.18a) in an asymptotic series in ζ^{-1} for ζ in the upper half ζ plane, upon comparing the two expansions, will give us the infinity of conservation laws.

From (2.18), we have

$$\ln a(\zeta)\simeq -i\frac{\epsilon^2m}{2\zeta\hbar^2}\left(\mathcal{N}-\frac{1}{2\zeta\hbar}\mathcal{P}+\frac{m}{2\zeta^2\hbar^2}\mathcal{E}+O(\zeta^{-3})\right), \tag{2.19}$$

where

$$\mathcal{N}\equiv\int_{-\infty}^{\infty}\Psi^*\Psi\,dx, \tag{2.20a}$$

$$\mathcal{P}\equiv-\tfrac{1}{2}i\hbar\int_{-\infty}^{\infty}(\Psi^*\Psi_x-\Psi_x^*\Psi)\,dx, \tag{2.20b}$$

$$\mathcal{E}\equiv\int_{-\infty}^{\infty}[(\hbar^2/2m)\Psi_x^*\Psi_x-\tfrac{1}{2}\epsilon^2(\Psi^*\Psi)^2]dx. \tag{2.20c}$$

Upon also expanding (2.17) and comparing with (2.19), we obtain

$$\mathcal{N}=\frac{2\hbar^2}{m\epsilon^2}\left(i\sum_{j=1}^{J}(\zeta_j^*-\zeta_j)+\frac{1}{2\pi}\int_{-\infty}^{\infty}d\xi\ln(1+\bar\rho\rho)\right), \tag{2.21a}$$

$$\mathcal{P}=\frac{2\hbar^3}{m\epsilon^2}\left(-i\sum_{j=1}^{J}(\zeta_j^{*2}-\zeta_j^2)-\frac{1}{\pi}\int_{-\infty}^{\infty}d\xi\,\xi\ln(1+\bar\rho\,\rho)\right), \tag{2.21b}$$

$$\mathcal{E}=\frac{4\hbar^4}{m^2\epsilon^2}\left(\frac{i}{3}\sum_{j=1}^{J}(\zeta_j^{*3}-\zeta_j^3)+\frac{1}{2\pi}\int_{-\infty}^{\infty}d\xi\,\xi^2\ln(1+\bar\rho\,\rho)\right). \tag{2.21c}$$

Thus by expanding $\ln a(\zeta)$, we can directly obtain the field number $\mathcal{N}$, the field momentum $\mathcal{P}$, and the field energy $\mathcal{E}$, in terms of the scattering data. As we can see from (2.21), these field variables (as well as the higher ones) only depend on $[\zeta_j]_{n=1}^{J}$ and $\bar\rho\rho$, which are the constants of the motion.

Before proceeding with quantization, let us stop for a moment and consider (2.20c) and (2.21c). By standard techniques, it is easy to show that (2.20c) is also the Hamiltonian for (2.3),[23] if we interpret the conjugate momentum of Ψ (Π) to be

$$\Pi=i\hbar\Psi^*. \tag{2.22}$$

The process of mapping the field (and its conjugate momentum $i\hbar\Psi^*$) into the scattering data S_+ is a canonical map. Therefore, (2.21c) is the Hamiltonian for the scattering data S_+, but which variables in the scattering data are the p's and which are the q's? To answer this, we must look at the simplicial form for this system,[19,20] which is

$$\begin{aligned}\int_{-\infty}^{\infty}\delta\Psi\wedge\delta\Pi\,dx=&\int_{-\infty}^{\infty}d\xi\,\delta(\arg b)\wedge\delta\left(\frac{\hbar^3}{\pi\epsilon^2m}\ln(1+\bar\rho\,\rho)\right)\\&+\sum_{j=1}^{J}\delta(\arg b_j)\wedge\delta\left(\frac{4\hbar^3}{\epsilon^2m}\eta_j\right)\\&+\sum_{j=1}^{J}\delta\left(\frac{\hbar^2}{\epsilon^2m}\ln(b_j^*b_j)\right)\wedge\delta(-2\hbar\,\xi_j),\end{aligned} \tag{2.23}$$

where

$$b_j=\rho_j\left.\frac{\partial a}{\partial\zeta}\right|_{\zeta=\zeta_j}\qquad(j=1,2,\ldots,J). \tag{2.24a}$$

$$\zeta_j=\xi_j+i\eta_j\quad(\eta_j>0) \tag{2.24b}$$

From (2.23), one can immediately read off the (p,q) variables, with the q's to the left of the "wedge" and the p's to the right.

3. QUANTIZATION

Using the ideas already laid out in the first two sections, we shall now identify our (p,q) conjugate variables for the scattering data. Upon making the simplest possible identifications, we first note that the system is constrained by certain natural nonholonomic constraints, which are constraints on the phase-space topology of the system. If we knew how to quantize in the presence of such constraints, there would be no problems, but it has already been shown[24] that the presence of such constraints can lead to inconsistent results upon quantization. In order to avoid this problem, we will first remain in the classical theory until we can find a classical canonical transformation which will allow these nonholonomic constraints to be reduced to trivial constraints. If this can be done (and for this field it can be done), then we may ignore these constraints and proceed with quantization.

When we have finally reduced all nonholonomic constraints to trivial constraints by classical canonical transformations, we then find that the Hamiltonian becomes a simple nonlinear function of harmonic oscillator and free particle Hamiltonians. Quantization therefore becomes almost trivial, and we can immediately give the eigenvalues of the number operator, field momentum operator, and the Hamiltonian. It is indeed

remarkable that these eigenvalues for the continuous part of the spectrum are identically the same as for the linear case ($\epsilon=0$). In other words, no effects requiring renormalization are found, and the zero point energy is independent of ϵ! When $\epsilon^2>0$, a discrete spectrum can also be present, and when it is, bound states occur, since the Hamiltonian is unbounded from below. Each bound state can be interpreted as being a bound state of n "excitations" moving as a coherent unit with a binding energy proportional to the cube of the number of excitations.

We shall now identify the (p,q) variables. Define η_j and ξ_j by (2.24b) and first take the p's $[A_j,\ p_j,\ p(\xi)]$ to be [see (2.23)]

$$A_j \equiv (4\hbar^3/m\epsilon^2)\eta_j \qquad (j=1,2,\ldots,J), \tag{3.1a}$$

$$p_j = -2\hbar\xi_j \tag{3.1b}$$

$$p(\xi) = (\hbar^3/\pi m\epsilon^2)\ln[1+\bar{\rho}(\xi)\rho(\xi)] \quad (\xi=\text{real}). \tag{3.1c}$$

Then from (2.23) the conjugate q's $[B_j,\ q_j,\ q(\xi)]$ are

$$B_j = \arg(b_j) \qquad (j=1,2,\ldots,J), \tag{3.2a}$$

$$q_j = +(\hbar^2/m\epsilon^2)\ln(b_j^* b_j) \tag{3.2b}$$

$$q(\xi) = \arg(b(\xi)) \quad (\xi=\text{real}). \tag{3.2c}$$

Before quantizing, we must look carefully at the topology of our phase space.[24] Since the Hamiltonian is separable, we need only consider the subspaces of a conjugate pair at a time. For the $(A_j,\ B_j)$ phase space, in order to be physical, we must demand

$$A_j > 0 \quad (j=1,2,\ldots,J) \tag{3.3}$$

since all eigenvalues must lie in the upper half ζ plane. This is a nonholonomic constraint and cannot be ignored. Also, we must demand B_j to be modulo 2π,

$$0 \le B_j < 2\pi, \tag{3.4}$$

which follows from (3.3a). We do not know how to quantize directly when a phase space has this topology. However, in this case, it turns out that we can perform a canonical transformation on this phase space and map it, one-to-one, onto the Cartesian phase space of the harmonic oscillator. This transformation is

$$A_j = \tfrac{1}{2}(P_j^2+Q_j^2), \tag{3.5a}$$

$$B_j = \arg(P_j + iQ_j), \tag{3.5b}$$

and in the (P_j, Q_j) phase space, the constraints (3.3), (3.4) are automatically and trivially satisfied.

Turning next to the (p_j, q_j) phase space, we see that it is already Cartesian, since there are no constraints on the real part of the eigenvalue or on the magnitude of ρ_j. Meanwhile, the topology of the phase space of $(p(\xi),\ q(\xi))$ is exactly the same as that of $(A_j,\ B_j)$, so that the same considerations apply.

Collecting everything together, we take our final (p,q) variables to be

$$\arg(P_j + iQ_j) = \arg(b_j), \tag{3.6a}$$

$$P_j^2 + Q_j^2 = (8\hbar^3/m\epsilon^2)\eta_j, \tag{3.6b}$$

$$p_j = -2\hbar\xi_j, \tag{3.6c}$$

$$q_j = (\hbar^2/m\epsilon^2)\ln(b_j^* b_j), \tag{3.6d}$$

$$b(\xi) = (P(\xi)+iQ(\xi))\,\tilde{R}(P^2(\xi)+Q^2(\xi)), \tag{3.6e}$$

where the function $\tilde{R}(Z)$ is defined by

$$\tilde{R}(Z) \equiv \left[Z^{-1}\exp\left(\frac{m\pi\epsilon^2}{2\hbar^3}Z\right) - Z^{-1}\right]^{1/2}. \tag{3.7}$$

Then, N, P, and $\mathcal{E}$ are given by

$$N = \sum_{j=1}^{J}\frac{1}{2\hbar}(P_j^2+Q_j^2) + \int_{-\infty}^{\infty} d\xi\,\frac{1}{2\hbar}[P^2(\xi)+Q^2(\xi)], \tag{3.8a}$$

$$P = \sum_{j=1}^{J}\frac{1}{2\hbar}(P_j^2+Q_j^2)p_j + \int_{-\infty}^{\infty} d\xi\,(-2\hbar\xi)\frac{1}{2\hbar}[P^2(\xi)+Q^2(\xi)], \tag{3.8b}$$

$$\mathcal{E} = \sum_{j=1}^{J}\frac{1}{2\hbar}(P_j^2+Q_j^2)\frac{p_j^2}{2m} - \frac{m\epsilon^4}{24\hbar^2}\sum_{j=1}^{J}\left(\frac{1}{2\hbar}(P_j^2+Q_j^2)\right)^3 + \int_{-\infty}^{\infty} d\xi\,\frac{(-2\hbar\xi)^2}{2m}\frac{1}{2\hbar}[P^2(\xi)+Q^2(\xi)]. \tag{3.8c}$$

Once we have the classical theory in the above form, it is easy to see how to interpret the various parts of the scattering data, as well as to see what the effects of quantization will be, since now we are only quantizing either independent harmonic oscillators or free particles. Upon quantization, we have

$$(1/2\hbar)(P_j^2+Q_j^2) \to \mathcal{A}_j^\dagger \mathcal{A}_j + \tfrac{1}{2}, \tag{3.9a}$$

$$(1/2\hbar)[P^2(\xi)+Q^2(\xi)] \to \mathcal{A}^\dagger(\xi)\mathcal{A}(\xi) + \tfrac{1}{2}, \tag{3.9b}$$

and $\mathcal{A}$ and $\mathcal{A}^\dagger$ are the usual destruction and creation operators, which satisfy

$$[\mathcal{A}_k, \mathcal{A}_j^\dagger] = \delta_j^k, \tag{3.9c}$$

$$[\mathcal{A}(\xi), \mathcal{A}^\dagger(\xi')] = \delta(\xi'-\xi). \tag{3.9d}$$

Defining the number operators by

$$N_j \equiv \mathcal{A}_j^\dagger \mathcal{A}_j, \tag{3.10a}$$

$$N(\xi) \equiv \mathcal{A}^\dagger(\xi)\mathcal{A}(\xi), \tag{3.10b}$$

then the operator form of (3.8) is simply

$$N_{\text{op}} = \sum_{j=1}^{J}(N_j+\tfrac{1}{2}) + \int_{-\infty}^{\infty} d\xi\,[N(\xi)+\tfrac{1}{2}], \tag{3.11a}$$

$$P_{\text{op}} = \sum_{j=1}^{J}(N_j+\tfrac{1}{2})(p_j)_{\text{op}} + \int_{-\infty}^{\infty} d\xi\,[N(\xi)+\tfrac{1}{2}]p(\xi), \tag{3.11b}$$

$$H = \mathcal{E}_{\text{op}} = \sum_{j=1}^{J}(N_j+\tfrac{1}{2})\frac{1}{2m}(p_j)_{\text{op}}^2 - \frac{m\epsilon^4}{24\hbar^2}\sum_{j=1}^{J}(N_j+\tfrac{1}{2})^3 + \int_{-\infty}^{\infty} d\xi\,[N(\xi)+\tfrac{1}{2}]\frac{1}{2m}p^2(\xi), \tag{3.11c}$$

where

$$p(\xi) \equiv -2\hbar\xi. \tag{3.12}$$

We see that (3.11) differs from the second quantized form of the linear Schrödinger equation (without a potential)[22] only by the soliton terms. Upon quantization, each soliton is comprised of n_j "excitations", where n_j

is an eigenvalue of N_j ($n_j = 0, 1, 2, \cdots$) and these n_j excitations form a bound state which moves at a momentum equal to an eigenvalue of $(p_j)_{op}$ ($-\infty < p_j < +\infty$) for a pure state. The energy of this bound state is the sum of a kinetic term and a "binding energy," which is proportional to $(n_j + \frac{1}{2})^3$ for a pure state. Thus the more excitations in a soliton, the more strongly they are bound. Note the conjugate variables and how they are paired. If we have an eigenstate of N_j, then, by the uncertainty principle, we know the amplitude and width of the soliton [see Eq. (2.15)], but we have no knowledge of the phase of the soliton. Similarly, if we know the momentum p_j of the soliton, we have no knowledge of the central position of the soliton.

So far, we have ignored the problem of factor ordering, and, based on the results for the N body problem with a δ-function interaction[16,17] and the appropriate limit of the quantized sine-Gordon equation,[12,25] it is known that the energy eigenvalues of the bound states should be $n_j^3 - n_j$ and not $(n_j + \frac{1}{2})^3$. Returning to (2.20c), one can see that the term $-n_j$ would come from the normal ordering of the nonlinear term in (2.20c), which would add a term proportional to (2.20a) to the Hamiltonian. But, at the present state, this would leave its coefficient in (3.8c) completely arbitrary. How may this be resolved? Consider the classical (or semiclassical) limit. By the inverse scattering transform (2.12), (2.13), we have a map from the classical limit of the operators $\mathcal{A}_j$ and $\mathcal{A}(\xi)$ to the fields $\Psi(x)$ and $\Psi^*(x)$. This map is nonlinear, and therefore, as we go to the full quantum domain, we must specify a factor ordering in order to have a unique map from the quantum operators $\mathcal{A}_j$ and $\mathcal{A}(\xi)$ to the operators $\Psi(x)$ and $\Psi^\dagger(x)$. Furthermore, since the commutation relation between Ψ and $\Psi^\dagger$ must be

$$[\Psi(x), \Psi^\dagger(y)] = \delta(x - y), \tag{3.13}$$

to all orders of $\hbar$ (by the correspondence principle, it is satisfied only to lowest order), we hypothesize that there exists a factor ordering for the inverse scattering transform which, due to (3.9c), (3.9d), will give (3.13). This condition essentially would uniquely specify the operator transform from scattering space to the field space, and therefore would uniquely determine the quantum form of (3.8c) which would correspond to the normal-ordered form of (2.20c).

ACKNOWLEDGMENTS

The author wishes to thank A. C. Newell of Clarkson College of Technology and D. W. McLaughlin and H. Flaska of the University of Arizona for their valuable comments and suggestions as to the action-angle variables, and also Mr. C.D. Levermore of NYU for his continual interest. In addition, the author wishes to thank R. Dashen, R. Jackiw, and L.D. Faddeev for their comments and suggestions.

*Research supported in part by NSF Grant #GP-43653.

[1] M.J. Ablowitz, D.J. Kaup, A.C. Newell, and H. Segur, Studies Appl. Math. **53**, 249 (1974).
[2] C.S. Gardner, J.M. Greene, M.D. Kruskal, and R. M. Miura, Phys. Rev. Lett. **19**, 1095 (1967); Comm. Pure Appl. Math. **37**, 97 (1974).
[3] V.E. Zakharov and A.B. Shabat, Zh. Eksp. Teor. Fiz. **61**, 118 (1971) [Sov. Phys. JETP **34**, 62 (1972)].
[4] M. Wadati, J. Phys. Soc. Japan **32**, 1681 (1972).
[5] M.J. Ablowitz, D.J. Kaup, A.C. Newell, and H. Segur, Phys. Rev. Lett. **30**, 1262 (1973).
[6] M.J. Ablowitz, D.J. Kaup, A.C. Newell, and H. Segur, Phys. Rev. Lett. **31**, 125 (1973).
[7] M.J. Ablowitz, D.J. Kaup, and A.C. Newell, J. Math. Phys. **15**, 1852 (1974).
[8] V.E. Zakharov, Zh. Eksp. Teor. Fiz. **65**, 219 (1973)[Sov. Phys. JETP **38**, 108 (1974)].
[9] V.E. Zakharov and S.V. Manakov, JETP Lett. **18**, 243 (1973).
[10] S.V. Manakov, Zh. Eksp. Teor. Fiz. **65**, 505 (1973).
[11] D.J. Kaup, Studies Appl. Math. **54**, 165 (1975).
[12] L.D. Faddeev and L.A. Takhtajan, "Essentially Nonlinear One-Dimensional Model of the Classical Field Theory" (preprint).
[13] D.J. Kaup "The Three-Wave Interaction, a Nondispersive Phenomenon" (to appear in Studies Appl. Math. in Dec., 1975).
[14] See, for example, F. Guerra, L. Rosen, and B. Simon, Commun. Math. Phys. **29**, 233 (1973), and references cited therein.
[15] A.C. Scott, F.Y.F. Chu, and D.W. McLaughlin, Proc. IEEE **61**, 1443 (1973).
[16] H. Bethe, Z. Physik **71**, 205 (1931).
[17] C.N. Yang, Phys. Rev. Lett. **19**, 1312 (1967); Phys. Rev. **168**, 1920 (1968).
[18] M.J. Ablowitz and A.C. Newell, J. Math. Phys. **14**, 1277 (1973).
[19] C.D. Levermore and D.J. Kaup (in preparation).
[20] H. Flaska and A.C. Newell, "Integrable Systems of Nonlinear Evolution Equations" (preprint).
[21] D.W. McLaughlin, J. Math. Phys. **16**, 96 (1975).
[22] V.E. Zakharov and L.D. Faddeev, Funct. Anal. Appl. 5, 280 (1971).
[23] See, for example, L.I. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1968), 3rd ed. Chap. 14.
[24] D.J. Kaup and A.P. Vitello, Phys. Rev. D **9**, 1648 (1974).
[25] R.F. Dashen, B. Hasslacher, and A. Neveu, "The Particle Spectrum in Model Field Theories from Semiclassical Functional Integral Techniques" (preprint).

Perturbative solution to order $\beta\epsilon$ of the Percus–Yevick equation for the square-well potential

Douglas Henderson

IBM Research Laboratory, San Jose, California 95193

Masao Chen

Department of Mathematics, Vanier College, Snowdon Campus, Montreal, Quebec, Canada
(Received 28 March 1975)

The radial distribution function (RDF) of a fluid is considered for the case of the square-well potential. If the RDF is expanded in powers of the depth of the square-well, ϵ, the first two terms are, in most applications, the most important. The Percus–Yevick (PY) integral equation for the RDF is examined and the resulting integral equations for these terms obtained. The first set of equations are just the PY equations for hard spheres which have been solved analytically. In this paper, the remaining equations for the terms of order $\beta\epsilon$, where $\beta = 1/k_BT$, are solved analytically and the results examined. We have speculated in the past that the PY theory could be used to obtain estimates of higher-order terms in a perturbation expansion of the RDF. We find that the PY theory cannot give reliable estimates of these higher-order terms for the square-well potential at high densities.

INTRODUCTION

In the statistical theory of fluids, a function of great interest is the *radial distribution function* (RDF), defined by

$$g(r_{12}) = V^2 \frac{\int \exp(-\beta\Phi)\, d\mathbf{r}_3 \cdots d\mathbf{r}_N}{\int \exp(-\beta\Phi)\, d\mathbf{r}_1 \cdots d\mathbf{r}_N}, \tag{1}$$

where $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ and $\mathbf{r}_i$ is the position of the ith molecule. In addition, V and Φ are, respectively, the volume and potential energy of the system of N molecules and $\beta = 1/k_BT$, where k_B is Boltzmann's constant and T is the temperature.

Not only does $g(r)$ give insight into the structure of the fluid, but the thermodynamic properties may be calculated from $g(r)$ by means of three distinct methods. First, there is the *compressibility equation*,

$$k_BT\left(\frac{\partial\rho}{\partial p}\right)_T = 1 + 4\pi\rho \int_0^\infty r^2 h(r)\, dr, \tag{2}$$

where $h(r) = g(r) - 1$ is the *total correlation function*, p is the pressure, and $\rho = N/V$, which can also be written in the completely equivalent form

$$\frac{1}{k_BT}\left(\frac{\partial p}{\partial\rho}\right)_T = 1 - 4\pi\rho \int_0^\infty r^2 c(r)\, dr, \tag{3}$$

where $c(r)$ is the *direct correlation function* which is defined by the Ornstein–Zernike equation:

$$h(12) = c(12) + \rho \int h(13)c(23)\, d\mathbf{r}_3, \tag{4}$$

where $h(12) = h(r_{12})$, etc.

The other two equations linking $g(r)$ and thermodynamics depend on the form of Φ. If the potential energy is made up of additive contributions of pairs of molecules, i.e.,

$$\Phi(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \sum_{i<j=1}^{N} u(r_{ij}), \tag{5}$$

the *pressure equation* has the form

$$\frac{pV}{Nk_BT} = 1 - \frac{2\pi\rho}{3k_BT} \int_0^\infty r^3 \frac{du}{dr} g(r)\, dr, \tag{6}$$

and the *energy equation* becomes

$$U = \tfrac{3}{2}Nk_BT + 2\pi N\rho \int_0^\infty r^2 u(r) g(r)\, dr. \tag{7}$$

In this paper we consider the square-well potential

$$u(r) = \begin{cases} \infty, & 0 < r < \sigma, \\ -\epsilon, & \sigma < r < \mu\sigma, \\ 0, & r > \mu\sigma. \end{cases} \tag{8}$$

Throughout this paper, we employ the usual choice, $\mu = \frac{3}{2}$. In the special case $\epsilon = 0$, or equivalently $T = \infty$, this potential becomes the hard-square potential. For the particular case of (8), the pressure equation may be simplified:

$$pV/NkT = 1 + 4\eta[xy(\sigma) + \mu^3(1-x)y(\mu\sigma)], \tag{9}$$

where $\eta = \pi\rho\sigma^3/6$ and $x = \exp(\beta\epsilon)$. The function $y(r)$, which is continuous even though $g(r)$ and $u(r)$ are not, is related to $g(r)$ by

$$g(r) = e(r)y(r), \tag{10}$$

where

$$e(r) = \exp[-\beta u(r)]. \tag{11}$$

The energy equation may be integrated to obtain the free energy A. For the particular case of the square-well potential

$$\frac{A}{Nk_BT} = \int_0^{\beta\epsilon} \frac{U}{N\epsilon}\, d(\beta\epsilon) = \frac{A_0}{Nk_BT} + 2\pi\rho \int_0^{\beta\epsilon} d(\beta\epsilon) \int_1^\mu r^2 g(r)\, dr, \tag{12}$$

where A_0 is the hard-sphere free energy. This quantity may be calculated from the expression of Carnahan and Starling[1]:

$$\frac{A_0}{Nk_BT} = \tfrac{3}{2}\ln\lambda - 1 + \ln\rho + \eta\frac{4-3\eta}{(1-\eta)^2}, \tag{13}$$

where $\lambda^2 = h^2/2\pi m k_BT$, m is the molecular mass and h is Planck's constant.

One of the most useful prescriptions for calculating $g(r)$ is that of Percus and Yevick[2] (PY). In the PY theory

$$c(r) = [e(r) - 1]y(r). \tag{14}$$

Elimination of $c(r)$ between (4) and (14) yields the PY integral equation:

$$y(12) = 1 + \rho \int [e(13)y(13) - 1][e(23) - 1]y(23)\, d\mathbf{r}_3. \tag{15}$$

We note that in the PY theory $c(r) = 0$ for $r > \mu\sigma$.

In this paper it is convenient to use, instead of (15), the completely equivalent integral equations of Baxter[3]:

$$rc(r) = -Q'(r) + 2\pi\rho \int_r^{\mu\sigma} Q'(t)Q(t-r)\, dt \tag{16}$$

and

$$rh(r) = -Q'(r) + 2\pi\rho \int_0^{\mu\sigma} (r-t)h(|r-t|)Q(t)\, dt, \tag{17}$$

where $Q'(r)$ is the derivative of $Q(r)$ and $Q(r) = 0$ for $r > \mu\sigma$. Thus,

$$Q(r) = -\int_r^{\mu\sigma} Q'(t)\, dt. \tag{18}$$

The function $Q(r)$ is continuous. Note, however, that $Q'(r)$ is not continuous. Baxter has also shown that the compressibility equation, Eqs. (2) or (3), is equivalent to

$$\frac{1}{k_B T}\left(\frac{\partial p}{\partial \rho}\right)_T = [\tilde{Q}(0)]^2, \tag{19}$$

where

$$\tilde{Q}(k) = 1 - 2\pi\rho \int_0^{\mu\sigma} \exp(ikr)Q(r)\, dr. \tag{20}$$

For an exact $g(r)$, Eqs. (2), (6), and (7) are equivalent. However, for an approximate theory, such as the PY theory, each of these equations yields somewhat different thermodynamic properties. In most cases, including the present one, (7) yields the most reliable results.

PERTURBATION EXPANSION

There is now a large body of evidence[4] which indicates that the properties of a fluid are dominated by the repulsive part of the potential. This idea actually dates back to van der Waals,[5] who assumed that the structure of a fluid is determined primarily by the repulsive forces and that the effect of the attractive forces is to provide a uniform background potential in which the molecules move.

This suggests that we might solve (16) and (17) by using the perturbation expansions

$$y(r) = y_0(r) + \beta\epsilon y_1(r) + \cdots, \tag{21}$$

$$c(r) = \begin{cases} -y_0(r) - \beta\epsilon y_1(r), & 0 < r < \sigma, \\ \beta\epsilon y_0(r), & \sigma < r < \mu\sigma, \\ 0, & r > \mu\sigma, \end{cases} \tag{22}$$

$$g(r) = \begin{cases} 0, & 0 < r < \sigma, \\ y_0(r) + \beta\epsilon[y_0(r) + y_1(r)], & \sigma < r < \mu\sigma, \\ y_0(r) + \beta\epsilon y_1(r), & r > \mu\sigma, \end{cases} \tag{23}$$

and

$$Q(r) = \begin{cases} Q_0(r) + \beta\epsilon Q_1(r), & 0 < r < \sigma, \\ \beta\epsilon Q_1(r), & \sigma < r < \mu\sigma, \\ 0, & r > \mu\sigma. \end{cases} \tag{24}$$

Thus, our objective is to obtain $Q_0(r)$ for $0 < r < \sigma$, $Q_1(r)$ for $0 < r < \mu\sigma$, and $y_0(r)$ and $y_1(r)$ for $r > 0$. To obtain the thermodynamic properties, it is only necessary, provided that $Q(r)$ is known, to know $y(r)$ for $\sigma \le r \le \mu\sigma$. Analytic expressions for $Q_0(r)$ and for $y_0(r)$ for $0 \le r \le 5\sigma$ have been obtained previously.[3,6-9] In this paper we obtain analytic expressions for $Q_1(r)$ and for $y_1(r)$ for $0 \le r \le 2\sigma$. Some of these results have been reported earlier.[10]

ZEROTH-ORDER RESULTS

If we let $s = r/\sigma$, we have from (16) and (17)

$$-sy_0(s) = -Q_0'(s) + 12\eta \int_s^1 Q_0'(t)Q_0(t-s)\, dt, \tag{25}$$

$$-s = -Q_0'(s) - 12\eta \int_0^1 (s-t)Q_0(t)\, dt \tag{26}$$

for $0 \le s \le 1$ and

$$sy_0(s) - s = 12\eta \int_0^{s-1} (s-t)y_0(s-t)Q_0(t)\, dt - 12\eta \int_0^1 (s-t)Q_0(t)\, dt \tag{27}$$

for $1 \le s \le 2$. In these equations, factors of σ^{-1} and σ^{-2} have been absorbed into the $Q_0'(s)$ and $Q_0(s)$, respectively.

Equations (25) and (26) have been solved by Baxter[3] who observed that (26) requires that

$$Q_0'(s) = as + b \tag{28}$$

and thus

$$Q_0(s) = \tfrac{1}{2}a(s^2 - 1) + b(s - 1). \tag{29}$$

Substitution of (28) and (29) into (26) yields

$$a = (1 + 2\eta)/(1 - \eta)^2 \tag{30}$$

and

$$b = -3\eta/2(1 - \eta)^2. \tag{31}$$

Substitution of these results into (25) yields

$$y_0(s) = a^2 - 6\eta(a + b)^2 s + \tfrac{1}{2}\eta a^2 s^3. \tag{32}$$

These results have been obtained earlier by Wertheim[6] and Thiele[7] by a direct solution of (15).

With these results, we may obtain the PY hard-sphere thermodynamic properties. From the pressure equation, we obtain

$$p_0 V/NkT = (1 + 2\eta + 3\eta^2)/(1 - \eta)^2 \tag{33}$$

and, from the compressibility equation, we obtain

$$p_0 V/NkT = (1 + \eta + \eta^2)/(1 - \eta)^3. \tag{34}$$

The energy equation result is obtained by differentiating A_0 in (12) and (13). This quantity is actually a constant of integration and so is not really a consequence of (7). In any case, the result is

$$p_0 V/NkT = (1 + \eta + \eta^2 - \eta^3)/(1 - \eta)^3. \tag{35}$$

We can solve (27) by letting $f(s) = sy_0(s)$ and changing variables in the first integral on the rhs. Thus,

$$f(s) = s + 12\eta \int_1^s f(t)Q_0(s-t)\, dt - 12\eta \int_0^1 (s-t)Q_0(t)\, dt. \tag{36}$$

We may solve (36) by differentiating three times to obtain the linear, homogeneous third-order ordinary dif-

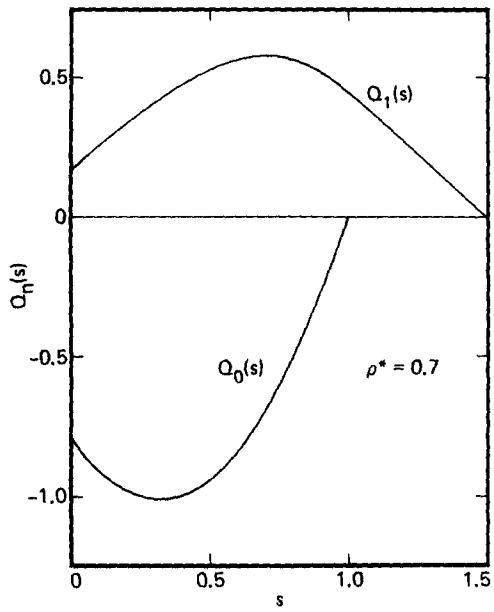


FIG. 1. $Q_0(s)$ and $Q_1(s)$ for PY theory for the square-well potential with $\mu = \frac{3}{2}$.

ferential equation

$$\mathcal{L}[f] = f'''(s) + \frac{6\eta}{1-\eta} f''(s) + \frac{18\eta^2}{(1-\eta)^2} f'(s) - \frac{12\eta(1+2\eta)}{(1-\eta)^2} f(s) = 0, \tag{37}$$

where $\mathcal{L}$ denotes the third-order linear differential operator.

If we seek a solution of the form

$$f(s) = A \exp(ms), \tag{38}$$

we find that

$$S(m) = m^3 + \frac{6\eta}{1-\eta} m^2 + \frac{18\eta^2}{(1-\eta)^2} m - \frac{12\eta(1+2\eta)}{(1-\eta)^2} = 0. \tag{39}$$

Thus, for $1 \leq s \leq 2$,

$$s y_0(s) = \sum_{i=0}^{2} A_i \exp(m_i s), \tag{40}$$

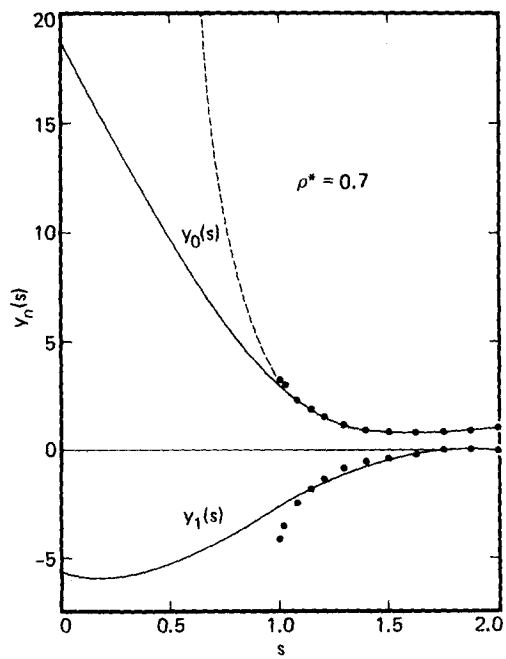


FIG. 2. $y_0(s)$ and $y_1(s)$ for the square-well potential with $\mu = \frac{3}{2}$. The points are MC values[11,13] and the solid and broken curves give the PY and Grundke—Henderson[12] values, respectively. The Grundke-Henderson values of $y_0(s)$ are obtained from a reliable parametrization.

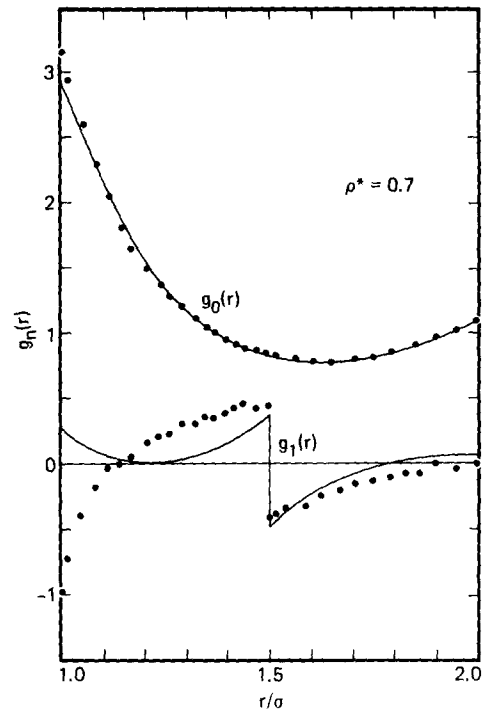


FIG. 3. $g_0(s)$ and $g_1(s)$ for the square-well potential with $\mu = \frac{3}{2}$. The points are MC values[11,13] and the curve gives the PY results.

where the m_i are the three roots of (39). Analytical expressions for the m_i are available.[6]

The A_i are determined from the boundary conditions

$$f(1) = 1 - 12\eta \int_0^1 Q_0(t)\,dt + 12\eta \int_0^1 t Q_0(t)\,dt, \tag{41}$$

$$f'(1) = 1 + 12\eta Q_0(0) f(1) - 12\eta \int_0^1 Q_0(t)\,dt, \tag{42}$$

and

$$f''(1) = 12\eta Q_0(0) f'(1) + 12\eta Q_0'(0) f(1). \tag{43}$$

Equations (41)—(43) are equivalent to the requirement that $y_0(r)$, $y_0'(r)$, and $y_0''(r)$ are continuous at $r = \sigma$.

Wertheim,[6] Chen, Henderson, and Davison,[8] and Smith and Henderson[9] have obtained these results earli-

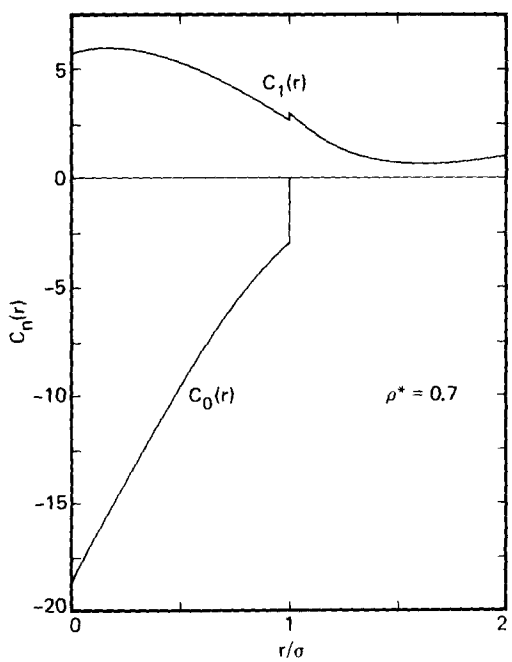


FIG. 4. $c_0(s)$ and $c_1(s)$ for PY theory for the square-well potential with $\mu = \frac{3}{2}$.

er by a different method. Smith and Henderson have obtained $y_0(r)$ analytically for $0 \le r \le 5\sigma$.

The functions $Q_0(s)$, $y_0(s)$, $g_0(s)$, and $c_0(s)$ are plotted in Figs. 1–4 for the relatively high density $\rho^* = \rho\sigma^3 = 0.7$. These results are typical of those obtained at other densities. Monte Carlo (MC) calculations, based on exact formulas for $g_0(s)$, and thus $y_0(s)$ for $s \ge 1$, are available.[11] In addition, there is the parametrization of $y_0(s)$ for $s \le 1$ of Grundke and Henderson[12] with which the PY $y_0(s)$ may be compared. Both the Monte Carlo and Grundke–Henderson results are plotted in Figs. 2 and 3. The PY $g_0(r)$ is quite good. As is well known, the most significant errors in the PY $g_0(s)$ are the slightly low values of $g_0(s)$ for $s \gtrsim 1$. On the other hand, the PY $y_0(s)$ is seriously in error for $s < 1$.

RESULTS FOR $Q_1(r)$

If we substitute (21)–(24) into (16) and (17), we obtain

$$-sy_1(s) = -Q_1'(s) + 12\eta \int_s^1 Q_0'(t)Q_1(t-s)\,dt + 12\eta \int_s^{s+1} Q_0(t-s)Q_1'(t)\,dt \tag{44}$$

and

$$Q_1'(s) = -12\eta \int_0^{3/2} (s-t)Q_1(t)\,dt + 12\eta \int_{s+1}^{3/2} (s-t)y_0(t-s)Q_1(t)\,dt \tag{45}$$

for $0 \le s \le \frac{1}{2}$,

$$-sy_1(s) = -Q_1'(s) + 12\eta \int_s^1 Q_0'(t)Q_1(t-s)\,dt + 12\eta \int_s^{3/2} Q_0(t-s)Q_1'(t)\,dt \tag{46}$$

and

$$Q_1'(s) = -12\eta \int_0^{3/2} (s-t)Q_1(t)\,dt \tag{47}$$

for $\frac{1}{2} \le s \le 1$, and

$$sy_0(s) = -Q_1'(s) + 12\eta \int_s^{3/2} Q_0(t-s)Q_1'(t)\,dt \tag{48}$$

and

$$sy_0(s) + sy_1(s) = -Q_1'(s) - 12\eta \int_0^{3/2} (s-t)Q_1(t)\,dt + 12\eta \int_0^{s-1} (s-t)[y_0(s-t) + y_1(s-t)]Q_0(t)\,dt + 12\eta \int_0^{s-1} (s-t)y_0(s-t)Q_1(t)\,dt \tag{49}$$

for $1 \le s \le \frac{3}{2}$. Note that we are explicitly using the value $\mu = \frac{3}{2}$.

Thus, to obtain $Q_1(r)$, we must solve these coupled integro-differential equations. The easiest equation with which to start is (48). If we set $f(s) = Q_1'(s)$ and $g(s) = sy_0(s)$, this equation becomes

$$f(s) = -g(s) + 12\eta \int_0^{3/2} Q_0(t-s)f(t)\,dt. \tag{50}$$

We solve (48) by differentiating three times to obtain

$$L'[f] = f'''(s) - \frac{6\eta}{1-\eta} f''(s) + \frac{18\eta^2}{(1-\eta)^2} f'(s) + \frac{12\eta(1+2\eta)}{(1-\eta)^2} f(s) = g'''(s). \tag{51}$$

Proceeding as before, we attempt a solution of the homogeneous portion of (51) which has the form

$$f(s) = B\exp(ms). \tag{52}$$

Thus,

$$D(m) = m^3 - \frac{6\eta}{1-\eta} m^2 + \frac{18\eta^2}{(1-\eta)^2} m + \frac{12\eta(1+2\eta)}{(1-\eta)^2} = 0. \tag{53}$$

We note that

$$D(-m) = -S(m). \tag{54}$$

Thus, the solution to the homogeneous equation is

$$f(s) = \sum_{l=0}^{2} B_l \exp(-m_l s), \tag{55}$$

where the m_l are the same as those in Eq. (40). Including the particular integral gives, for $1 \le s \le \frac{3}{2}$,

$$Q_1'(s) = \sum_{l=0}^{2} [B_l \exp(-m_l s) + C_l \exp(m_l s)], \tag{56}$$

where

$$C_l = -A_l m_l^3 / D(m_l). \tag{57}$$

The B_l are determined from the boundary conditions

$$f(\tfrac{3}{2}) = -g(\tfrac{3}{2}), \tag{58}$$

$$f'(\tfrac{3}{2}) = -g'(\tfrac{3}{2}) - 12\eta Q_0(0) f(\tfrac{3}{2}), \tag{59}$$

and

$$f''(\tfrac{3}{2}) = -g''(\tfrac{3}{2}) - 12\eta Q_0(0) f'(\tfrac{3}{2}) + 12\eta Q_0'(0) f(\tfrac{3}{2}). \tag{60}$$

Integrating (56) yields, for $1 \le s \le \frac{3}{2}$,

$$Q_1(s) = -\sum_{l=0}^{2} [B_l \exp(-m_l s) - C_l \exp(m_l s)]/m_l + C, \tag{61}$$

where C is determined from the boundary condition

$$Q_1(\tfrac{3}{2}) = 0. \tag{62}$$

Next we consider Eqs. (45) and (47). From these we obtain

$$Q_1'(s) = 12\eta[-As + B - \psi(s)] \tag{63}$$

for $0 \le s \le 1$, where

$$A = \int_0^{3/2} Q_1(t)\,dt, \tag{64}$$

$$B = \int_0^{3/2} tQ_1(t)\,dt, \tag{65}$$

and

$$\psi(s) = \int_{s+1}^{3/2} (t-s)y_0(t-s)Q_1(t)\,dt. \tag{66}$$

Substituting (40) and (61) into (66) gives

$$\psi(s) = a + \sum_{l=0}^{2} b_l \exp(-m_l s) + s\sum_{l=0}^{2} c_l \exp(-m_l s) + \sum_{l=0}^{2} d_l \exp(m_l s) \tag{67}$$

for $0 \le s \le \frac{1}{2}$, and

$$\psi(s) = 0 \tag{68}$$

for $\frac{1}{2} \le s \le 1$. We note that because of (68), $Q_1'(s)$ is continuous at $s = \frac{1}{2}$. In (67)

$$a = -C\sum_{l=0}^{2} \frac{A_l \exp(m_l)}{m_l}, \tag{69}$$

$$b_l = -\frac{A_l B_l}{2m_l} + C\frac{A_l}{m_l} \exp\left(\frac{3m_l}{2}\right) + \sum_{k=0}^{2} \frac{A_l C_k}{m_k(m_l + m_k)} \exp[\tfrac{3}{2}(m_l + m_k)]$$

$$- \sum_{k\neq l=0}^{2}\left(\frac{A_l B_k}{m_k(m_l - m_k)} \exp[\tfrac{3}{2}(m_l - m_k)]\right.$$

$$\left. - \frac{A_k B_l}{m_l(m_k - m_l)} \exp(m_k - m_l)\right), \tag{70}$$

$$c_l = \frac{A_l B_l}{m_l}, \tag{71}$$

and

$$d_l = - \sum_{k=0}^{2} \frac{A_k C_l}{m_l(m_l + m_k)} \exp(m_l + m_k). \tag{72}$$

Hence, integrating (63) gives

$$Q_1(s) = 12\eta[- \tfrac{1}{2}As^2 + Bs + D - \phi(s)] \tag{73}$$

for $0 < s < \frac{1}{2}$, where D is a constant of integration, and

$$\phi(s) = as - \sum_{l=0}^{2}\left(\frac{b_l}{m_l} - \frac{c_l}{m_l^2}\right)\exp(- m_l s)$$

$$- s\sum_{l=0}^{2} \frac{c_l}{m_l} \exp(- m_l s) + \sum_{l=0}^{2} \frac{d_l}{m_l} \exp(m_l s) \tag{74}$$

and

$$Q_1(s) = 12\eta(- \tfrac{1}{2}As^2 + Bs + E) \tag{75}$$

for $\frac{1}{2} < s < 1$, where E is a constant of integration. We note that $Q_0(s)$ is also of the form of (75). It is easy to prove that all $Q_n(s)$, and thus $Q(s)$, are quadratic functions of s for $\mu - 1 \leq s \leq 1$.

We determine D and E by requiring that $Q_1(s)$ be continuous at $s = \frac{1}{2}$ and $s = 1$ and we determine A and B from (64) and (65). The function $Q_1(s)$ is now known.

We have plotted $Q_1(s)$ for $\rho^* = 0.7$ in Fig. 1. As already pointed out, there is no discontinuity in the slope of $Q_1(s)$ at $s = \frac{1}{2}$. Although it is not seen in Fig. 1, $Q_1(s)$ has a small discontinuity in slope at $s = 1$.

RESULTS FOR $y_1(s)$

Now let us consider Eq. (49). Let us set $f(s) = sy_1(s)$ and $g(s) = sy_0(s)$. Thus, for $1 \leq s \leq \frac{3}{2}$,

$$f(s) = - g(s) - Q_1'(s) - 12\eta \int_0^{3/2} (s - t)Q_1(t)\,dt$$

$$+ 12\eta \int_1^s [g(t) + f(t)]Q_0(s - t)\,dt$$

$$+ 12\eta \int_1^s g(t)Q_1(s - t)\,dt. \tag{76}$$

We may solve (76) by differentiating three times to obtain

$$\mathcal{L}[f] = - Q_1^{iv}(s) + 12\eta[Q_1(0)g''(s) + Q_1'(0)g'(s) + Q_1''(0)g(s)$$

$$+ 12\eta \int_1^s g(t)Q_1'''(s - t)\,dt, \tag{77}$$

where the linear operator $\mathcal{L}$ is given by (37). In obtaining (77) we have used the fact that $\mathcal{L}[g] = 0$.

Hence

$$\mathcal{L}[f] = P(s), \tag{78}$$

where

$$P(s) = \sum_{l=0}^{2} X_l \exp(- m_l s) + \sum_{l=0}^{2} Y_l \exp(m_l s) + R(s), \tag{79}$$

$$X_l = m_l^3 B_l, \tag{80}$$

$$Y_l = 12\eta[Q_1(0)m_l^2 + Q_1'(0)m_l + Q_1''(0)]A_l - m_l^3 C_l, \tag{81}$$

and

$$R(s) = 12\eta \int_1^s g(t)Q_1'''(s - t)\,dt. \tag{82}$$

Now for $0 \leq t \leq \frac{1}{2}$

$$Q_1'''(t) = \sum_{l=0}^{2} p_l \exp(- m_l t) + s\sum_{l=0}^{2} q_l \exp(- m_l t)$$

$$+ \sum_{l=0}^{2} r_l \exp(m_l t), \tag{83}$$

where

$$p_l = 12\eta b_l m_l^2 - 24\eta c_l m_l, \tag{84}$$

$$q_l = - 12\eta c_l m_l^2, \tag{85}$$

and

$$r_l = - 12\eta d_l m_l^2. \tag{86}$$

Substitution of (40) and (83) into (82) gives

$$P(s) = \sum_{l=0}^{2} (P_l + sQ_l) \exp(- m_l s) + \sum_{l=0}^{2} (R_l + sT_l) \exp(m_l s), \tag{87}$$

where

$$P_l = X_l + 12\eta(q_l - p_l)\sum_{k=0}^{2} \frac{A_k}{(m_l + m_k)} \exp(m_l + m_k)$$

$$- 12\eta q_l \sum_{k=0}^{2} \frac{A_k}{(m_l + m_k)^2} \exp(m_l + m_k), \tag{88}$$

$$Q_l = - 12\eta q_l \sum_{k=0}^{2} \frac{A_k}{(m_l + m_k)} \exp(m_l + m_k), \tag{89}$$

$$R_l = Y_l + 12\eta A_l \sum_{k=0}^{2}\left(\frac{p_k}{(m_l + m_k)} + \frac{q_k}{(m_l + m_k)^2}\right)$$

$$+ 12\eta A_l \sum_{k\neq l=0}^{2} \frac{r_k}{(m_l - m_k)} - 12\eta r_l \sum_{k\neq l=0}^{2} \frac{A_k}{(m_k - m_l)}$$

$$\times \exp(m_k - m_l) - 12\eta A_l r_l, \tag{90}$$

and

$$T_l = 12\eta A_l r_l. \tag{91}$$

The solution of the homogeneous portion of (78) is

$$f(s) = \sum_{l=0}^{2} \alpha_l \exp(m_l s). \tag{92}$$

Including the particular integral gives for $1 \leq s \leq \frac{3}{2}$

$$sy_1(s) = \sum_{l=0}^{2} (\alpha_l + \beta_l s + \gamma_l s^2) \exp(m_l s)$$

$$+ \sum_{l=0}^{2} (\delta_l + s\epsilon_l) \exp(- m_l s), \tag{93}$$

where

$$\gamma_l = \frac{T_l}{2S'(m_l)}, \tag{94}$$

$$\beta_l = \frac{R_l - \gamma_l S''(m_l)}{S'(m_l)}, \tag{95}$$

$$\epsilon_l = - \frac{Q_l}{D(m_l)}, \tag{96}$$

$$\delta_l = \frac{- P_l + \epsilon_l D'(m_l)}{D(m_l)}, \tag{97}$$

and $S(m)$ and $D(m)$ are given by (39) and (53), respec-

tively. The α_l are determined from the boundary conditions

$$f(1) = -g(1) - Q_1'(1) - 12\eta(A - B), \tag{98}$$

$$f'(1) = -g'(1) - Q_1''(1) - 12\eta A + 12\eta\{[g(1) + f(1)]Q_0(0) + g(1)Q_1(0)\}, \tag{99}$$

and

$$f''(1) = -g''(1) - Q_1'''(1) + 12\eta\{[g'(1) + f'(1)]Q_0(0) + g'(1)Q_1(0) + [g(1) + f(1)]Q_0'(0) + g(1)Q_1'(0)\}. \tag{100}$$

Equations (98)—(100) ensure the continuity of $y_1(s)$ and its first two derivatives at $s = 1$.

To obtain $y_1(s)$ for $\frac{3}{2} \leq s \leq 2$, we must solve

$$\begin{aligned} sy_1(s) = {} & 12\eta \int_0^{s-3/2} (s-t)y_1(s-t)Q_0(t)\,dt \\ & + 12\eta \int_{s-3/2}^{s-1} (s-t)[y_0(s-t) + y_1(s-t)]Q_0(t)\,dt \\ & + 12\eta \int_0^{s-1} (s-t)y_0(s-t)Q_1(t)\,dt \\ & + 12\eta \int_0^{3/2} (s-t)Q_1(t)\,dt. \end{aligned} \tag{101}$$

Let us set $f(s) = sy_1(s)$ for $\frac{3}{2} \leq s \leq 2$, $h(s) = sy_1(s)$ for $1 \leq s \leq \frac{3}{2}$, and $g(s) = sy_0(s)$ for $1 \leq s \leq 2$. Hence,

$$\begin{aligned} f(s) = {} & 12\eta \int_{3/2}^{s} f(t)Q_0(s-t)\,dt + 12\eta \int_1^{3/2} [h(t) + g(t)]Q_0(s-t)\,dt \\ & + 12\eta \int_1^{s} g(t)Q_1(s-t)\,dt - 12\eta \int_0^{3/2} (s-t)Q_1(t)\,dt. \end{aligned} \tag{102}$$

If we differentiate three times,

$$\begin{aligned} \mathcal{L}[f] = {} & 12\eta[Q_1(0)g''(s) + Q_1'(0)g'(s) + Q_1''(0)g(s)] \\ & + 12\eta \int_1^{s} g(t)Q_1'''(s-t)\,dt, \end{aligned} \tag{103}$$

where the linear operator $\mathcal{L}$ is given by (37). Thus,

$$\mathcal{L}[f] = T(s), \tag{104}$$

where

$$T(s) = \sum_{l=0}^{2} Z_l \exp(m_l s), \tag{105}$$

$$\begin{aligned} Z_l = {} & 12\eta A_l \sum_{k=0}^{2} \frac{p_k}{m_l + m_k}\left[1 - \exp\left(-\frac{m_l + m_k}{2}\right)\right] \\ & + \sum_{k=0}^{2} \frac{q_k}{(m_l + m_k)^2}\left[1 - \exp\left(-\frac{m_k + m_l}{2}\right)\right] \\ & - \sum_{k=0}^{2} \frac{q_k}{2(m_l + m_k)} \exp\left(-\frac{m_l + m_k}{2}\right) \\ & - \sum_{k \neq l=0}^{2} \frac{r_k}{m_k - m_l}\left[1 - \exp\left(\frac{m_k - m_l}{2}\right)\right] + \tfrac{1}{2} r_l \\ & + 12\eta[Q_1(0)m_l^2 + Q_1'(0)m_l + Q''(0)]A_l. \end{aligned} \tag{106}$$

Therefore, the solution of the homogeneous equation is

$$f(s) = \sum_{l=0}^{2} \theta_l \exp(m_l s). \tag{107}$$

Including the particular integral gives for $\frac{3}{2} \leq s \leq 2$

$$sy_1(s) = \sum_{l=0}^{2} (\theta_l + s\phi_l)\exp(m_l s), \tag{108}$$

where

$$\phi_l = Z_l / S'(m_l) \tag{109}$$

and $S(m)$ is given by (39). The θ_l are determined from the boundary conditions

$$f(\tfrac{3}{2}) = h(\tfrac{3}{2}), \tag{110}$$

$$f'(\tfrac{3}{2}) = h'(\tfrac{3}{2}) + g'(\tfrac{3}{2}) + 12\eta Q_0(0)f(\tfrac{3}{2}) + Q_1''(\tfrac{3}{2}) - 12\eta Q_0(0)[h(\tfrac{3}{2}) + g(\tfrac{3}{2})] \tag{111}$$

and

$$f''(\tfrac{3}{2}) = h''(\tfrac{3}{2}) + g''(\tfrac{3}{2}) + 12\eta Q_0(0)f'(\tfrac{3}{2}) + 12\eta Q_0'(0)f(\tfrac{3}{2}) + Q_1'''(\tfrac{3}{2}) - 12\eta Q_0(0)[h'(\tfrac{3}{2}) + g'(\tfrac{3}{2})] - 12\eta Q_0'(0)[h(\tfrac{3}{2}) + g(\tfrac{3}{2})]. \tag{112}$$

Equations (110)—(112) ensure the continuity of $y_1(s)$ and its first and second derivatives at $s = \frac{3}{2}$.

Finally, we have (44) and (46) which yield $y_1(s)$ for $0 \leq s \leq 1$. In contrast to the other three equations (45), (47), and (101), these equations can be integrated straightforwardly. Inasmuch as the thermodynamic properties do not depend on $y_1(s)$ for $0 \leq s \leq 1$, these integrations were performed numerically.

Results for $y_1(s)$, $g_1(s)$, and $c_1(s)$ are plotted in Figs. 2—4 for $\rho^* = 0.7$. It is seen that $g_1(s)$ is smaller than $g_0(s)$ whereas $y_1(s)$ is actually bigger than $y_0(s)$ for $s \sim 1$. As a result, we might expect the perturbation expansion to converge more rapidly for $g(s)$ than for $y(s)$. Indeed, if $y(r)$ were calculated from a first-order perturbation expansion, negative values of $y(r)$ would be obtained for high densities and low temperatures.

Comparison with the Monte Carlo[13] calculations of $g_1(s)$ and $y_1(s)$ shows that the PY values of $g_1(s)$ and $y_1(s)$ are most seriously in error for s less than 1.1σ. The errors in the PY $g_1(s)$ and $y_1(s)$ are much larger than was the case for $g_0(s)$ and $y_0(s)$.

THERMODYNAMIC PROPERTIES

If we substitute (23) into (12), we obtain the perturbation expansion energy equation of state:

$$A = A_0 + \beta\epsilon A_1 + (\beta\epsilon)^2 A_2 + \cdots, \tag{113}$$

where A_0 is given by Eq. (13),

$$\begin{aligned} \frac{A_1}{Nk_BT} = {} & -12\eta \int_1^{3/2} s^2 y_0(s)\,ds \\ = {} & -12\eta \sum_{l=0}^{2} \frac{A_l}{m_l^2}\left[(\tfrac{3}{2}m_l - 1)\exp(\tfrac{3}{2}m_l) - (m_l - 1)\exp(m_l)\right] \end{aligned} \tag{114}$$

and

$$\begin{aligned} A_2/Nk_BT = {} & -6\eta \int_1^{3/2} s^2[y_0(s) + y_1(s)]\,ds \\ = {} & \tfrac{1}{2}(A_1/Nk_BT) - 6\eta \sum_{l=0}^{2} \{(\alpha_l/m_l^2)[(\tfrac{3}{2}m_l - 1)\exp(\tfrac{3}{2}m_l) \\ & - (m_l - 1)\exp(m_l)] - (\beta_l/m_l^3)[(\tfrac{9}{4}m_l^2 - 3m_l + 2) \\ & \times \exp(\tfrac{3}{2}m_l) - (m_l^2 - 2m_l + 2)\exp(m_l)] \\ & - (\gamma_l/m_l^4)[(\tfrac{27}{8}m_l^3 - \tfrac{27}{4}m_l^2 + 9m_l - 6)\exp(\tfrac{3}{2}m_l) \\ & - (m_l^3 - 3m_l^2 + 6m_l - 6)\exp(m_l)] \\ & + (\delta_l/m_l^2)[-(\tfrac{3}{2}m_l + 1)\exp(\tfrac{3}{2}m_l) + (m_l + 1)\exp(m_l)] \\ & + (\epsilon_l/m_l^3)[-(\tfrac{9}{4}m_l^2 + 3m_l - 2)\exp(\tfrac{3}{2}m_l) \\ & + (m_l^2 + 2m_l - 2)\exp(m_l)]\}. \end{aligned} \tag{115}$$

The pressure is obtained by differentiating:

$$\frac{p_n}{\rho k_B T} = \rho \frac{dA_n/NkT}{d\rho} . \tag{116}$$

If we substitute (23) into (9), we obtain the perturbation expansion of the pressure equation of state

$$p = p_0 + \beta\epsilon p_1 + \cdots , \tag{117}$$

where p_0 is given by Eq. (33) and

$$\begin{aligned} p_1/k_B T &= -4\eta[\tfrac{27}{8} y_0(\tfrac{3}{2}) - y_0(1) - y_1(1)] \\ &= -4\eta \sum_{i=0}^{2} [\tfrac{9}{4} A_i \exp(\tfrac{3}{2} m_i) - (A_i + \alpha_i + \beta_i + \gamma_i) \\ &\quad \times \exp(m_i) - (\delta_i + \epsilon_i) \exp(-m_i)] . \end{aligned} \tag{118}$$

If we substitute (24) into (19), we obtain the perturbation expansion of the compressibility equation of state

$$\left(\frac{\partial p}{\partial \rho}\right)_T = \left(\frac{\partial p_0}{\partial \rho}\right)_T + \beta\epsilon \left(\frac{\partial p_1}{\partial \rho}\right)_T + \cdots , \tag{119}$$

where p_0 is given by Eq. (34),

$$\frac{1}{k_B T}\left(\frac{\partial p_1}{\partial \rho}\right)_T = 2\tilde{Q}_0(0)\tilde{Q}_1(0), \tag{120}$$

$$\tilde{Q}_0(0) = \frac{1 + 2\eta}{(1-\eta)^2} , \tag{121}$$

and

$$\begin{aligned} \tilde{Q}_1(0) &= -12\eta \int_0^{3/2} Q_1(s)\, ds \\ &= -12\eta \sum_{i=0}^{2} \{(B_i/m_i)[\tfrac{3}{2}\exp(-\tfrac{3}{2}m_i) - \exp(-m_i)] \\ &\quad - (C_i/m_i)[(\tfrac{1}{2}\exp(\tfrac{3}{2}m_i) - \exp(m_i)]\}. \end{aligned} \tag{122}$$

It is to be noted that in the energy equation of state A_n (or p_n) is a functional of $y_{n-1}(r)$ whereas in the pressure and compressibility equations of state p_n is a functional of $y_n(r)$. Thus, the energy equation of state gives one higher-order term in the perturbation expansion than does either the pressure or compressibility equations of state. In addition, the energy equation involves an integral of $g(r)$ over values of r for which the PY $g(r)$ seems fairly reliable. In contrast, the pressure and compressibility equations involve values of $y(r)$, where the PY theory is less reliable. As a result, the energy equation of state may be expected to be more reliable than either the pressure or compressibility equation of state.

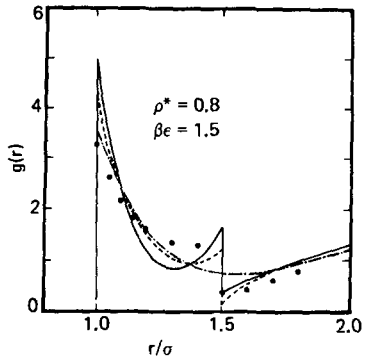


FIG. 6. RDF for square-well potential for $\mu = \frac{3}{2}$. The points are MC values values[15] and the curves marked — — —, — — — —, and —— give the PY values for $g_0(s)$, $g_0(s) + \beta\epsilon g_1(s)$, and $g(s)$, respectively.

Values for the thermodynamic functions have been published earlier.[10] With the exception of A_2, these values are correct. However, in the course of this work we discovered an error in our computer program for A_2. For this reason we show A_2 in Fig. 5. At low densities, where the PY theory is accurate, the agreement with the Monte Carlo values of A_2[13] is good. However, at high densities the PY values of A_2 are too small in absolute magnitude.

CONCLUDING REMARKS

We have seen that the PY $g_0(r)$ is fairly good, although too small at contact. Also, the PY $g_1(r)$ is rather poor especially near contact. One might expect that good results could be obtained by correcting the errors in the PY $g(r)$. That is, $g(r)$ might be calculated from

$$g(r) = g_0(r) + \beta\epsilon g_1(r) + [g^{PY}(r) - g_0^{PY}(r) - \beta\epsilon g_1^{PY}(r)]. \tag{123}$$

Actually, this was the motivation for this work.

We have compared the PY $g(r)$ for the square-well potential[14] with the MC values[15] for $\rho^* = 0.8$ and $\beta\epsilon = 1.5$ in Fig. 6. It is seen that the PY $g(r)$ is too large for $r \gtrsim 1$ and $r > 1.4$ and too small in between. For comparison we have plotted the PY values for $g_0(r)$ and $g_0(r) + \beta\epsilon g_1(r)$ in Fig. 6. In Fig. 7 we have compared,

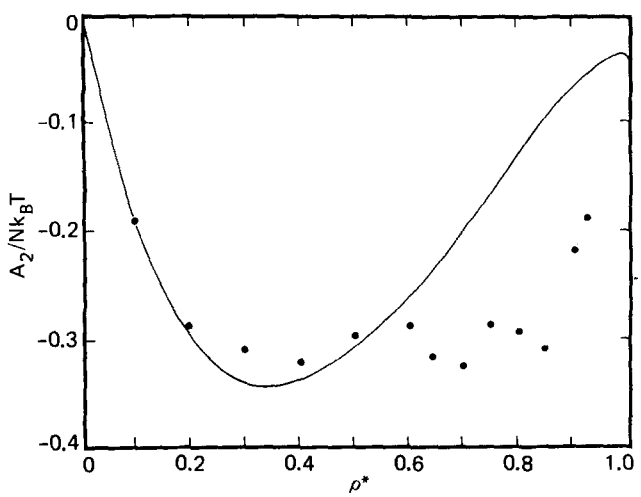


FIG. 5. A_2 for the square-well potential with $\mu = \frac{3}{2}$. The points are MC values[13] and the curve gives the PY results.

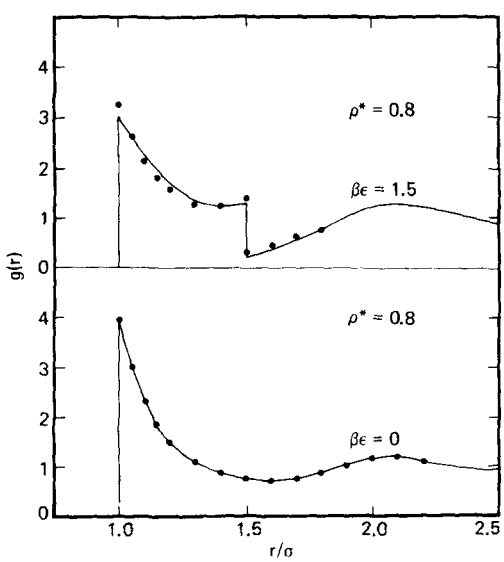


FIG. 7. RDF for square-well potential for $\mu = \frac{3}{2}$. The points are MC values[15] and the curves are calculated from first-order perturbation theory using MC values values[11,13] for $g_0(s)$ and $g_1(s)$.

for $\rho^* = 0.8$, the MC values[11,13] of $g_0(r)$ and $g_0(r) + \beta\epsilon g_1(r)$ with the MC values[14] of $g(r)$. The agreement is quite good. In fact, it is much better than that obtained from (123). The PY estimate of the contribution of the higher-order terms is qualitatively satisfactory inasmuch as it is positive for $r \gtrsim 1$ and $r > 1.4$ and negative for $1.2 \lesssim r \lesssim 1.4$. However, the PY estimate is much too large in magnitude. Thus, the PY theory cannot be used to obtain estimates of higher-order perturbation terms at high densities for the square-well potential.

This may be seen from the thermodynamic properties also. The pressure at the high density $\rho^* = 0.85$ is plotted in Fig. 6 of Ref. 14. The use of Eq. (123) corrects the initial values and slopes of the pressure and compressibility isochores. Even if this is done, the pressure and compressibility isochores remain unsatisfactory at low temperatures. On the other hand, the use of Eq. (123) with the energy equation does yield excellent values for the pressure. However, this is somewhat fortuitous because the energies and, more especially, the heat capacities obtained in this manner show appreciable errors.

We conclude that for the square-well potential the higher-order perturbation terms are not adequately described by the Percus–Yevick theory.

One might also use Eq. (123) with the g's replaced by y's. However, this is even less satisfactory.

[1]N.F. Carnahan and K.E. Starling, J. Chem. Phys. **51**, 635 (1969).
[2]J.K. Percus and G.J. Yevick, Phys. Rev. **110**, 1 (1958).
[3]R.J. Baxter, Austr. J. Phys. **21**, 563 (1968).
[4]J.A. Barker and D. Henderson, Ann. Rev. Phys. Chem. **23**, 439 (1972).
[5]J.D. van der Waals, thesis, Leiden (1873).
[6]M. Wertheim, Phys. Rev. Lett. **10**, 321 (1963); J. Math. Phys. **5**, 643 (1964).
[7]E. Thiele, J. Chem. Phys. **39**, 474 (1963).
[8]R. Chen, D.Henderson, and S.G. Davison, Proc. Natl. Acad. Sci. USA **54**, 1514 (1965).
[9]W.R. Smith and D. Henderson, Mol. Phys. **19**, 411 (1970).
[10]M. Chen, D. Henderson, and J.A. Barker, Canadian J. Phys Phys. **47**, 2009 (1969).
[11]J.A. Barker and D. Henderson, Mol. Phys. **21**, 187 (1971).
[12]E.W. Grundke and D. Henderson, Mol. Phys. **24**, 269 (1972).
[13]W.R. Smith, D. Henderson, and J.A. Barker, J. Chem. Phys. **55**, 4027 (1971).
[14]W.R. Smith, D. Henderson, and R.D. Murphy, J. Chem. Phys. **61**, 2911 (1974).
[15]D. Henderson, unpublished work.

Internal-labeling operators*

R. T. Sharp†

Centre de Recherches Mathématiques, Université de Montréal, Montréal 101, Canada
(Received 16 May 1975)

A simple method is described for finding all possible "missing label" operators when a semisimple group is reduced according to a maximal semisimple subgroup. The operators may be chosen to be Hermitian and hence lead to an orthonormal basis. The solution is worked out for all seven cases of one missing label. In each case two independent subgroup scalars are found in the enveloping algebra of the group; either of them can be used as the missing label.

1. INTRODUCTION

In applications of compact groups to physical problems, it is often necessary to decompose the irreducible representations (IR's) of a group into IR's of a subgroup. It frequently happens that the subgroup of physical interest does not provide enough labels to specify the basis states unambiguously.

Elliott[1] solved this problem for the group-subgroup $SU(3)\supset O(3)$ by a projection technique; good O(3) states are projected from certain "intrinsic" states, and the missing label is provided by the intrinsic state from which the projection is made. The same approach has been used for $O(5)\supset SU(2)\times U(1)^2$ and for $SU(4)\supset SU(2)\times SU(2)$.[3] A different solution of the internal labeling problem is provided by the method of elementary multiplets (elementary permissible diagrams).[4,5,6,7] It has been shown[8] that all subgroup IR's of all group IR's are defined by the stretched products of powers of a finite number of elementary multiplets (subgroup IR's belonging to low-lying group IR's); the exponents of the powers of the elementary multiplets supply the missing (as well as nonmissing) labels. The two methods described in this paragraph define analytic but generally nonorthonormal bases.

This paper is concerned with a complementary approach to the missing label problem, namely, the use of basis functions which are the common eigenstates of some complete set of commuting Hermitian operators. Besides the Casimir operators of group and subgroup and appropriate internal subgroup operators an additional number[9] $\frac{1}{2}(r_G - l_G - r_H - l_H)$ of labeling operators must be found; r_G, l_G, r_H, l_H are the order and rank of group and subgroup. These missing label operators should be subgroup scalars in the enveloping algebra of the group.

For $SU(3)\supset O(3)$ there is one missing label. The SU(3) generators consist of an O(3) vector L_i [the O(3) generators] and an O(3) second rank tensor Q_j. Bargmann and Moshinsky[4] suggested that the missing label could be

$$X^{(3)}=\sum_{ij} L_i Q_j L_{-i-j}\begin{pmatrix}1 & 2 & 1\\ i & j & -i-j\end{pmatrix}.$$

Later Racah[10] stated without proof that the most general missing label is a polynomial in $X^{(3)}$ and a second operator

$$X^{(4)}=\sum_{ijk} L_i Q_j Q_{k-j} L_{-k-i}\begin{pmatrix}1 & 1 & 2\\ i & -k-i & k\end{pmatrix}$$

$$\times\begin{pmatrix}2 & 2 & 2\\ j & k-j & -k\end{pmatrix}(-1)^k$$

(actually the missing label may also depend on the group and subgroup Casimir operators and internal subgroup labels; we will not repeat this obvious point in similar situations throughout this paper).

A proof of Racah's conjecture is given by Judd, Miller, Patera and Winternitz.[11] These authors show how for any group–subgroup a generating function can be constructed for subgroup scalars in the enveloping algebra of the group. Each labeling operator is characterized by its degrees in the various subgroup tensors occurring in the adjoint representation of the group. Thus for $SU(3)\supset O(3)$ the generating function turns out to be

$$\frac{1+L^3Q^3}{(1-L^2)(1-Q^2)(1-Q^3)(1-L^2Q)(1-L^2Q^2)}.$$

Each factor in the denominator defines an independent labeling operator. Thus L^2, Q^2, Q^3 are the O(3) and SU(3) Casimir operators [actually the second order SU(3) Casimir operator is a linear combination of Q^2 and L^2; the third order SU(3) Casimir operator is a combination of Q^3 and L^2Q. We do not belabor this point in similar situations later in this paper]. L^2Q and L^2Q^2 are the labeling operators of Racah. The fact that L^3Q^3 occurs only linearly implies that its square is a polynomial in the other operators and hence redundant. The derivation treats the tensors L and Q as c-numbers; but in fact they are operators and L^3Q^3 itself can be expressed in terms of the commutator of L^2Q and L^2Q^2, so should be dropped as a member of the integrity basis of subgroup scalars. Although this method solves in principle the problem of determining all possible labeling operators for any group–subgroup, it turns out that the determination of the generating function is extremely laborious in practice, even in the simplest cases.

In Sec. 2 we describe a method for obtaining much more simply the information contained in the generating function. In Sec. 3 the method is applied to determine the most general labeling operator for all cases (seven) where there is one missing label. Section 4 contains a discussion of difficulties which arise when there is more than one missing label and possible future research.

2. DESCRIPTION OF THE METHOD

We are faced with the problem of finding all subgroup scalars in the enveloping algebra of a group. More pre-

cisely we wish to find a finite integrity basis, a set of elementary subgroup scalars in terms of which all can be expressed as polynomials.

To this end we divide the elements of the group algebra into two sets of subgroup tensors. If we can now enumerate all higher tensors which can be formed as polynomials in the tensors of each set, we can also enumerate all subgroup scalars—each corresponds to a pair of tensors, one from each set, which transform by conjugate IR's of the subgroup. In this way we learn just the information given by the generating function of Judd *et al.*,[11] and can straightforwardly determine the finite integrity basis for subgroup scalars.

We illustrate the method by considering a case in which the subgroup is O(3) and the group generators decompose into two O(3) tensors one of which is L whose components are the O(3) generators, the other of which is an O(3) tensor Q of integer rank j. The integrity basis for higher tensors formed from L consists of L^2 and L. Setting aside L^2, we note that there is just one way to construct a tensor of integer rank k from the components of L, and it is of degree k. It follows that there is one-to-one correspondence between O(3) tensors constructed as polynomials in the components of Q alone and O(3) scalars in the enveloping algebra of the group.

To find the finite integrity basis for tensors which are polynomials in Q we consider the equivalent problem of finding the elementary multiplets for the labeling problem $SU(2j+1) \supset O(3)$ restricted to the symmetric IR's $(\lambda 0 \cdots 0)$ of $SU(2j+1)$; or if we set aside the O(3) scalar Q^2 the problem becomes that of finding elementary multiplets for the labeling problem $O(2j+1) \supset O(3)$ restricted to the IR's $(0 \cdots 0\lambda)$ of $O(2j+1)$. From each elementary multiplet thus found we form an O(3) scalar by contracting with the tensor of the same rank formed from powers of L. Since the counting is identical for the two problems it follows that the scalars thus formed, together with L^2 and Q^2, are an integrity basis for O(3) scalars in the enveloping algebra of the group.

The same procedure can be used, *mutatis mutandi*, for each group–subgroup of interest.

3. CASES WITH ONE MISSING LABEL

In this section the methods of the preceding section are used to determine the most general operator which can be used to solve the labeling problem for each of the seven cases (hopefully we have not overlooked any) with one missing label and a maximal subgroup. First we remark that when the method is used for a case with no missing label the integrity basis which one finds consists simply of the Casimir operators of group and subgroup.

A. SU(3) ⊃ O(3)

The group generators decompose into two O(3) tensors, the O(3) generators L and a rank 2 tensor Q. Setting aside L^2 and Q^2, we observe that the members of the integrity basis for O(3) scalars correspond to the elementary multiplets for the labeling problem O(5) ⊃ O(3) restricted to the IR's (0λ) or O(5). But these elementary multiplets are known[8] to be (1,2), (2,2), (3,0), (3,3) with $(3,3)^2$ redundant; the notation is (λ, l). It follows that the integrity basis for O(3) scalars in the enveloping algebra of SU(3) consists of L^2, Q^2, L^2Q, L^2Q^2, Q^3, L^3Q^3 with $(L^3Q^3)^2$ redundant, in agreement with the result of Judd *et al.*[11] Since this case has already been discussed we pass on to the next.

B. O(5) ⊃ SU(2) × U(1)

The O(5) generators decompose under SU(2)×U(1) into an SU(2)×U(1) scalar D and three SU(2) vectors for which the U(1) label D has the values $\pm 1, 0$. The components of the $D=0$ vector L are the SU(2) generators. Let us denote by U, V the vectors with $D=\pm 1$. We break the generators into two sets, one consisting of L, the other of U and V, setting aside for the moment the SU(2)×U(1) scalars D, L^2, $UV = \sum U_i V_i (-1)^i$. Then the integrity basis for SU(2) tensors with $D=0$ formed from the components of U and V consists of the elementary multiplets for the labeling problem $SU(3) \supset O(3) \sim SU(2)$ restricted to the self-conjugate IR's (λ, λ) of SU(3). The elementary multiplets for this case are easily shown to be (1,1), (1,2), (2,0), (2,2) with $(2,2)^2$ redundant [the notation is (λ, l)]. It follows that the finite integrity basis for SU(2)×U(1) scalars in the enveloping algebra of O(5) consists of D, L^2, UV, UVL, UVL^2, U^2V^2, $U^2V^2L^2$ with $(U^2V^2L^2)^2$ redundant. Now D, L^2, UV, U^2V^2 correspond to the U(1), SU(2), and O(5) Casimir operators. There are two independent missing label operators, namely

$$UVL = \sum_{ij} U_i L_j V_{-i-j} \begin{pmatrix} 1 & 1 & 1 \\ i & j & -i-j \end{pmatrix}$$

$$UVL^2 = \sum_{ijk} L_i U_j V_{k-j} L_{-k-i} \begin{pmatrix} 1 & 1 & 2 \\ i & -k & -i\ k \end{pmatrix} \times \begin{pmatrix} 1 & 1 & 2 \\ j & k-j & -k \end{pmatrix} (-1)^k.$$

$U^2V^2L^2$ should be dropped from the integrity basis since it is expressible in terms of the commutator of UVL and UVL^2. The most general missing label is a polynomial in UVL and UVL^2.

C. G_2 ⊃ SU(3)

The G_2 generators decompose under SU(3) into a (1,1) octet T [the SU(3) generators], a (10) triplet U, and a (01) triplet V. We put U and V in one set, T in the other.

Leaving aside for the moment the SU(3) scalar UV, we note that there is just one way to construct the SU(3) IR (pq) as a stretched product of powers of U and V—it is of degree p in U and q in V.

Turning to the octet T and leaving aside for the time being the SU(3) scalar T^2 we note that the integrity basis for SU(3) tensors which are polynomials in the components of T consists of the elementary multiplets for the labeling problem O(8) ⊃ SU(3) restricted to the IR's (000λ) of O(8). These elementary multiplets are found to be (1,11), (2,11), (3,03), (3,30), (3,00) with $(2,11)^3$ redundant.

It follows that the finite integrity basis for SU(3) scalars in the enveloping algebra of G_2 consists of T^2, UV, UVT, UVT^2, U^3T^3, V^3T^3, T^3 with $(UVT^2)^3$ redun-

dant. Actually, when operator properties are taken into account, $(UVT^2)^2$ is redundant since it can be expressed in terms of the commutator of U^3T^3 and UVT. The SU(3) scalars T^2, T^3, UV, and V^3T^3 can be identified with the Casimir operators of $SU(3)$ and G_2. Thus the most general missing label operator is a polynomial in UVT and U^3V^3 (and the Casimir operators) plus another polynomial multiplied by UVT^2.

D. SU(4) > SU(2) × SU(2) × U(1)

The SU(4) generators consist of D, an SU(2)×SU(2) ×U(1) scalar, the generators S and T of the two SU(2) groups, each with the U(1) label $D=0$ and two $(\frac{1}{2},\frac{1}{2})$ tensors U and V with $D=\pm 1$, respectively. We put S and T in one set of tensors, U and V in the other, leaving aside for now the SU(2)×SU(2)×U(1) scalars D, S^2, T^2. Then our SU(2)×SU(2) tensor (s,t), s, t integer, can be formed in just one way from the components of S and T.

Turning to U and V, we notice that the finite integrity basis for SU(2)×SU(2) tensors with $D=0$ which are polynomials in U and V consists of the scalar UV and the elementary multiplets for the labeling problem SU(4) ⊃ SU(2)×SU(2) restricted to SU(4) IR's $(\lambda 0 \lambda)$; here SU(2)×SU(2) is included in SU(4) as in the Wigner supermultiplet model. The elementary multiplets are found to be[6] (1,11), (1,10), (1,01), (2,00), (2,11) with $(2,11)^2$ redundant.

It follows that the integrity basis for SU(2)×SU(2) ×U(1) scalars in the enveloping algebra of SU(4) consists of D, S^2, T^2, UV, UVS, UVT, $UVST$, U^2V^2, U^2V^2ST with $(U^2V^2ST)^2$ redundant. D, S^2, T^2, UV, UVS, U^2V^2 represent the U(1), SU(2)×SU(2), SU(4) Casimir operators. The independent missing label operators are

$$UVT=\sum_{ijk} U_{ij}T_kV_{-i,-j-k}\begin{pmatrix}1&1&1\\ j&k&-j-k\end{pmatrix}(-1)^i,$$

$$UVST=\sum_{ijlm} S_iU_{jl}V_{-i-j,m}T_{-l-m}\begin{pmatrix}1&1&1\\ i&j&-i-j\end{pmatrix}\begin{pmatrix}1&1&1\\ l&m&-l-m\end{pmatrix}.$$

U^2V^2ST is redundant, since it is expressible in terms of the commutator of UVT and $UVST$. The most general missing label is a polynomial in UVT, $UVST$.

E. Sp(6) ⊃ Sp(4) × SU(2)

The generators of Sp(6) consist of an SU(2) vector L [the SU(2) generators], an Sp(4) (20) tensor Q [the SP(4) generators] and a tensor T which is a (10) tensor under Sp(4) and an $l=\frac{1}{2}$ spinor under SU(2). We put L and Q in one set of Sp(4)×SU(2) tensors and T in the other.

First we find the integrity basis for Sp(4)×SU(2) tensors which are polynomials in the components of T. They are the elementary multiplets for the labeling problem SU(8) ⊃ Sp(4)×SU(2) restricted to symmetric SU(8) IR's $(\lambda 000000)$. These elementary multiplets turn out to be $(1;10,\frac{1}{2})$, $(2;01,0)$, $(2;00,0)$ [the notation in $(\lambda;pq,l)$]. Leaving aside for the time being the Sp(4) ×SU(2) scalar $T^2 \sim (2;00,0)$, there is just one way to construct an Sp(4) tensor of rank (p,q) from the components of T. It will have rank $\frac{1}{2}p$ as an SU(2) tensor.

Next let us find the integrity basis for Sp(4) tensors formed from the components of Q. This mean finding the elementary multiplets for the labeling problem O(10) ⊃ Sp(4) restricted to the IR's (0000λ) of O(10) [we are leaving aside for the moment the Sp(4) scalar Q^2]. These elementary multiplets turn out to be (1,20), (2,01), (2,02), (3,20), (4,00), (4,21) with $(4,21)^2$ redundant [the notation is (λ,pq)].

Thus we find that the integrity basis for Sp(4)×SU(2) scalars in the enveloping algebra of Sp(6) consists of L^2, Q^2, Q^4, T^2, Q^2T^2, Q^2T^4, QT^2L, Q^3T^2L, Q^4T^4L with $(Q^4T^4L)^2$ redundant. The first six operators listed above correspond respectively to the SU(2), Sp(4), and Sp(6) Casimir operators, QT^2L and Q^3T^2L are independent missing label operators, in terms of which the general missing label operator is a polynomial. Q^4T^4L should be dropped, for it is expressible in terms of the commutator of QT^2L and Q^3T^2L.

F. O(7) ⊃ G_2

The generators of O(7) decompose under G_2 into a (01) tensor T (the G_2 generators) and a (10) tensor S.

To find the integrity basis for G_2 tensors formed from the components of S we need the elementary multiplets for the labeling problem O(7) ⊃ G_2, restricted to the O(7) IR's (00λ) (we are setting aside for the moment the G_2 scalar S^2). The solution of this problem is known.[7] There is just a single elementary multiplet (1,10).

The integrity basis for G_2 tensors formed from T consists of the elementary multiplets for the labeling problem O(14) ⊃ G_2 restricted to the O(14) IR's (000000λ) (we ignore for now the G_2 scalar T^2). We need to consider only those elementary multiplets with second G_2 IR label zero, since the rest cannot combine with stretched powers of the (10) tensor S to form a scalar. The relevant elementary multiplets are (2,20), (3,10), (3,30), (4,20), (6,00), (6,30) with $(6,30)^2$ redundant.

Hence the integrity basis for G_2 scalars in the enveloping algebra of O(7) consists of T^2, T^6, S^2, T^2S^2, T^3S^3, T^3S, T^4S^2, T^6S^3 with $(T^6S^3)^2$ redundant. T^2, T^6 are the G_2 Casimir operators; S^2, T^2S^2, T^3S^3 represent those of O(7). Thus T^3S or T^4S^2 or a polynomial in them can serve as the missing label. T^6S^3 can be expressed in terms of the commutator of T^3S and T^4S^2 and should be dropped.

G. SU(3) × SU(3) ⊃ SU(3)

This is the external labeling problem for SU(3). O'Raifeartaigh[12] suggested a mixed or polarized Casimir operator to lift the degeneracy in the SU(3) external labeling problem. His operator is a linear combination of UV^2 and U^2V below; its matrix elements with respect to a certain nonorthonormal basis were given by Chew and Sharp.[13] We will see that this is not the most general missing label.

The SU(3)×SU(3) generators decompose into two octets U and V under SU(3). The integrity basis for tensors formed from either is given in subsection C of this section. It is straightforward to show that the integrity basis for scalars formed from these tensors consists of

U^2, V^2, UV, U^2V, UV^2, U^2V^2, U^3, V^3, U^3V^3 with $(U^3V^3)^2$ redundant. U^2, V^2, UV and U^3, V^3, U^2V represent the quadratic and cubic Casimir operators, respectively, of the three SU(3) groups. UV^2, U^2V^2 or any polynomial in them is the most general missing label. U^3V^3 is dropped because it is expressible in terms of the commutator of UV^2 and U^2V^2.

4. DISCUSSION

The forms of the general missing label operators for the seven cases with one missing label show a remarkable similarity. In each case there are two independent operators [and, for $G_2 \supset SU(3)$, a third which can appear at most linearly because its square is expressible in terms of the others] and the missing operator is an arbitrary polynomial in these two. The freedom permitted in choosing the missing label should be used to make it correspond, perhaps empirically, to some quantity of physical interest.

Several difficulties arise when one attempts to treat group—subgroup problems with two or more missing labels.

The determination of the relevant integrity basis for subgroup scalars becomes more difficult. The details have been worked out, however, for $SU(4) \supset SU(2) \times SU(2)$ (Wigner supermultiplet model) by Miller,[14] using the generating function method; his results have been checked by the method of the present paper. The work has also been done for the group—subgroup $O(5) \supset SU(2)$. However no attempt has yet been made in these two cases to reduce the rather large number of scalars in the integrity basis by utilizing relations between them arising out of commutation rules.

A more fundamental problem to be solved is that of determining the most general *pair* of *commuting* missing labels. Moshinsky and Nagel[15] have given one example of such a commuting pair for the case $SU(4) \supset SU(2) \times SU(2)$.

As pointed out by Racah,[10] the eigenvalues and eigenstates of any operator of the type derived in the article must be determined numerically in practice, when the subgroup multiplicity exceeds two. Such calculations have already been carried out extensively for the group—subgroup $SU(3) \supset O(3)$.[11,16] Similar calculations would be useful for many of the group-subgroup pairs discussed above in Sec. 3.

ACKNOWLEDGMENTS

The author is grateful for helpful discussions during the preparation of this paper with J. Patera, M. Perroud, and P. Winternitz. The branching tables of Patera and Sankoff[17] were of great assistance in determining the elementary multiplets for the various labeling problems considered.

*Supported by the National Research Council of Canada and the Ministère de l'Education du Québec.

†On leave from the Department of Physics, McGill University, Montreal, Quebec, Canada.

[1]J. P. Elliott, Proc. Roy. Soc. Lond. A **245**, 128, 562 (1958).
[2]K. Ahmed and R. T. Sharp, J. Math. Phys. **11**, 1112 (1970).
[3]J. P. Draayer, J. Math. Phys. **11**, 3225 (1970); K. Ahmed and R. T. Sharp, Ann. Phys. **71**, 421 (1972).
[4]V. Bargmann and M. Moshinsky, Nucl. Phys. **23**, 177 (1961).
[5]V. Syamela Devi and T. Venkatarayudu, J. Math. Phys. **9**, 1057 (1968); V. Syamela Devi and M. Moshinsky, J. Math. Phys. **10**, 455 (1969); V. Syamela Devi, Nuovo Cimento A **68**, 1 (1970).
[6]R. T. Sharp and C. S. Lam, J. Math. Phys. **10**, 2033 (1969).
[7]B. G. Wybourne, J. Math. Phys. **13**, 455 (1972).
[8]R. Weitzenböck, Acta Math. **58**, 231 (1932).
[9]G. Racah, Lecture notes, "Group Theory and Spectroscopy," Institute for Advanced Study, Princeton, N.J., 1951; reprinted in Ergeb. Exakt. Naturw. **37**, 28 (1965).
[10]G. Racah, in *Group Theoretical Concepts and Methods in Elementary Particles Physics*, edited by F. Gürsey (Gordon and Breach, New York, 1964).
[11]B. R. Judd, W. Miller, Jr., J. Patera, and P. Winternitz, J. Math. Phys. **15**, 1787 (1974).
[12]L. O'Raifeartaigh, "Simple Solution to the Multiplicity Problem in Higher Symmetries," a talk given at the Toronto Conference on Symmetries and Particle Physics, 1965.
[13]C. K. Chew and R. T. Sharp, Can. J. Phys. **44**, 2789 (1966).
[14]W. Miller, Jr., private communication.
[15]M. Moshinsky and J. G. Nagel, Phys. Lett. **5**, 173 (1963).
[16]McKay, J. Patera, and R. T. Sharp (to be published); M. Moshinsky, J. Patera, R. T. Sharp and P. Winternitz (to be published).
[17]J. Patera and D. Sankoff, *Tables of Branching Rules for Representations of Simple Lie Algebras* (Presses de l'Université de Montréal, Montréal, 1973).

An asymptotic formula for the twisted product of distributions

Kuang Chi Liu*

Department of Mathematics, State University of New York College at Oswego, Oswego, New York 13126
(Received 11 February 1975)

The expansion theorem for the twisted product associated with the Weyl form of CCR for n degrees of freedom is generalized to involve tempered distributions.

I. INTRODUCTION

In a previous paper by the author,[1] a mathematically rigorous expansion theorem for the twisted product associated with the Weyl form of CCR for n degrees of freedom was obtained. The resulting asymptotic formula is valid for the class of c^∞ functions rapidly decreasing at ∞. In this paper, we extend our previous results to the twisted product of g and T, where g is a c^∞ function on the phase space R^{2n}, slowly decreasing at ∞, and T, any tempered distribution.

II. NOTATIONS AND PRELIMINARIES

Let $\alpha=(\alpha_1, \alpha_2, \ldots, \alpha_k)$ be a k-tuple index of nonnegative integers. We shall write

$$\alpha! = \alpha_1!\alpha_2! \cdots \alpha_k!,$$

$$|\alpha| = \sum_{j=1}^{k} \alpha_j.$$

For $x=(x_1, x_2, \ldots, x_k)\in R^k$

$$x^\alpha = x_1^{\alpha_1}x_2^{\alpha_2}\cdots x_k^{\alpha_k},$$

$$D^\alpha = (-i)^{|\alpha|}\left(\frac{\partial}{\partial x_1}\right)^{\alpha_1}\left(\frac{\partial}{\partial x_2}\right)^{\alpha_2}\cdots\left(\frac{\partial}{\partial x_k}\right)^{\alpha_k};$$

when $k=2n$,

$$\partial^\alpha = (-i)^{|\alpha|}\left(\frac{\partial}{\partial x_{n+1}}\right)^{\alpha_1}\left(\frac{\partial}{\partial x_{n+2}}\right)^{\alpha_2}\cdots\left(\frac{\partial}{\partial x_n}\right)^{\alpha_n} \times\left(-\frac{\partial}{\partial x_1}\right)^{\alpha_{n+1}}\left(-\frac{\partial}{\partial x_2}\right)^{\alpha_{n+2}}\cdots\left(-\frac{\partial}{\partial x_n}\right)^{\alpha_{2n}}.$$

Let $S(R^k)$ denote the space of c^∞ complex functions on R^k, rapidly decreasing at ∞. For $\Phi\in S(R^k)$, define its Fourier transform $F\Phi$ by

$$(F\Phi)(\xi) = (2\pi)^{-k/2}\int \exp(i\xi\cdot x)\Phi(x)\,dx,$$

and the inverse of F is given by

$$(F^{-1}\Phi)(x) = (2\pi)^{-k/2}\int \exp(-ix\cdot\xi)\,\Phi(\xi)\,d\xi,$$

where $\cdot$ is the inner product in R^k. Unless otherwise indicated, our integrals are over R^k with $d\xi$ denoting the Lebesque measure. $S(R^k)$ can be given a Hausdorff locally convex topology which is metrizable and complete. The topological dual of $S(R^k)$, denoted by $S'(R^k)$, is known as the space of tempered distributions. Let $T\in S'(R^k)$, we write

$\langle T, \Phi\rangle$ to denote the value of T on $\Phi\in S(R^k)$,

and its Fourier transform and the inverse, also denoted by F, F^{-1}, are defined by

$$\langle FT, \Phi\rangle = \langle T, F\Phi\rangle,$$

$$\langle F^{-1}T, \Phi\rangle = \langle T, F^{-1}\Phi\rangle \quad \forall\ \Phi\in S(R^k).$$

It is well known that F, F^{-1} are continuous [in the sense of strong topology of $S'(R^k)$] isomorphisms from $S'(R^k)$ onto itself.

Finally, we denote by $O_M(R^k)$ the space of c^∞ functions on R^k, slowly increasing at ∞ and also $O_c'(R^k)$, the space of distributions rapidly decreasing at ∞. It is known that elements in $O_M(R^k)$ and $O_c'(R^k)$ are tempered and $S(R^k)$ is contained in both. It is also true that F, F^{-1} are topological isomorphisms from $O_M(R^k)$ onto $O_c'(R^k)$ and $O_c'(R^k)$ onto $O_M(R^k)$. For details about tempered distributions, see Ref. 2.

III. TWISTED CONVOLUTION AND TWISTED PRODUCT

In the following, we extend the definitions of the twisted convolution and the twisted product to involve tempered distributions. Let us recall that if f, $g\in S(R^{2n})$, the convolution of f and g is

$$(f * g)(\eta) = \int f(\xi)g(\eta-\xi)\,d\xi,$$

and the twisted convolution of f and g of index $c>0$ is[1]

$$(f * cg)(\eta) = \int \exp[-ic(\eta\times\xi)/2]\,f(\xi)g(\eta-\xi)\,d\xi,$$

where $\eta=(u,v)$, $\xi=(s,t)$ are elements of R^{2n} and $\eta\times\xi = u\cdot t - v\cdot s$.

Let $E_\eta(c,\xi)=\exp[ic(\eta\times\xi)/2]$, we have the following:

Lemma 3.1:

$$(f * cg)(\eta) = f*(E_\eta g)(\eta),$$

$$f*cg\in S(R^{2n}) \quad \forall\ f, g\in S(R^{2n}).$$

Proof: Since $E_\eta g\in S(R^{2n})$ for each $\eta\in R^{2n}$, we have

$$f*(E_\eta g)(\eta) = \int f(\xi)(E_\eta g)(\eta-\xi)\,d\xi = \int f(\xi)\exp[ic\eta\times(\eta-\xi)/2]g(\eta-\xi)\,d\xi.$$

Now $\eta\times(\eta-\xi) = -\eta\times\xi$; therefore,

$$f*(E_\eta g)(\eta) = \int \exp[-ic(\eta\times\xi)/2]\,f(\xi)g(\eta-\xi)\,d\xi = (f*cg)(\eta).$$

$f*cg\in S(R^{2n})$ follows from the fact that convolution is a closed operation in $S(R^{2n})$. | | |

If one considers $f*g\in S(R^{2n})$ as a tempered distribution, one has, after a change of variable,

$$\langle f*g, \Phi\rangle = \int\int f(\xi)g(\eta)\,\Phi(\xi+\eta)\,d\xi\,d\eta = \int f(\xi)\langle g_\eta, \Phi(\xi+\eta)\rangle\,d\xi$$

$$= \langle f_\xi, \langle g_\eta, \Phi(\xi+\eta)\rangle\rangle \quad \forall\ \Phi \in S(R^{2n}).$$

Similarly, considering $f * cg$ as a tempered distribution, one has

$$\langle f * cg, \Phi\rangle = \langle f * (E_\eta g), \Phi\rangle$$
$$= \langle f_\xi, \langle (E_\xi g)_\eta, \Phi(\xi+\eta)\rangle\rangle \quad \forall\ \Phi \in S(R^{2n}).$$

Note that $f * cg \neq g * cf$ in general.

Motivated by the above, we make the following definitions:

Definition 3.1: Let $f \in O_c'(R^{2n})$ and $T \in S'(R^{2n})$. The twisted convolution of f and T (of index c) is given by

$f * cT = f * (E_\xi(c)T)$ in the following sense:

$$\langle f * cT, \Phi\rangle = \langle f * (E_\xi(c)T), \Phi\rangle$$
$$= \langle f_\xi, \langle (E_\xi(c)T)_\eta, \Phi(\xi+\eta)\rangle\rangle \quad \forall\ \Phi \in S(R^{2n}).$$

In the above definition, for each $\xi \in R^{2n}$ and $c > 0$, $E_\xi(c)T$ is the tempered distribution given by

$$\langle (E_\xi(c)T)_\eta, \Phi(\xi+\eta)\rangle = \langle T_\eta, E_\xi(c)\Phi(\xi+\eta)\rangle = \Psi(\xi)$$

and $\Psi \in S(R^{2n})$. It follows from the property of convolution[2] that $f * cT = f * (E_\xi(c)T)$ is again a tempered distribution.

Definition 3.2: Let $g \in O_M(R^{2n})$ and $T \in S'(R^{2n})$. The twisted product (of index c) $g \circ cT$ is given by,

$$g \circ cT = (2\pi)^{-n} F^{-1}[(Fg) * c(FT)].$$

Since $Fg \in O_c'(R^{2n})$ and $FT \in S'(R^{2n})$ so that $(Fg) * c(FT) \in S'(R^{2n})$, therefore $g \circ cT$ is well defined and is again a tempered distribution.

Remark 3.1: Definition 3.2 is designed so that the Weyl operator preserves the twisted product. If f, g are elements of $S(R^{2n})$, the Weyl operator $w(f \circ cg)$ corresponding to $f \circ cg$ has the property that $w(f \circ cg) = w(f)w(g)$. For details, see Ref. 1.

IV. AN ASYMPTOTIC FORMULA FOR $g \circ cT$

The following is a generalized version of the expansion theorem in Ref. 1, valid for c^∞ functions on R^{2n}, slowly increasing at ∞.

Theorem 4.1: Let $g \in O_M(R^{2n})$ and $T \in S'(R^{2n})$ and

$$G(c, \Phi) = \langle g \circ cT, \Phi\rangle, \quad \Phi \in S(R^{2n}).$$

Then, for $c > 0$ and a positive integer N,

$$G(c, \Phi) = \sum_{k=0}^{N} \langle b_k[g, T], \Phi\rangle c^k + o(c^N)$$

as $c \downarrow 0$ for each Φ, where

$$b_k[g, T] = \left(\frac{i}{2}\right)^k \sum \frac{1}{\alpha!} D^\alpha g \partial^\alpha T, \quad k \geq 0,$$

is a tempered distribution.

Remark 4.1: The summation in b_k is over all distinct $2n$-tuples α, such that $|\alpha| = k$. $D^\alpha g$ and $\partial^\alpha T$ are to be understood in the distribution sense.

The following lemmas are needed to give a proof to Theorem 4.1.

Lemma 4.1: Considered as a function of c, $G(c, \Phi)$ is infinitely differentiable for each $\Phi \in S(R^{2n})$, and

$$\frac{\partial^k G(c, \Phi)}{\partial c^k} = (2\pi)^{-n}\left(\frac{i}{2}\right)^k \cdot \langle (Fg)_\eta, \langle (FT)_\xi, (\eta \times \xi)^k E_\eta(c) \times F^{-1}\Phi(\eta+\xi)\rangle\rangle.$$

Proof: For each $\Phi \in S(R^{2n})$

$$G(c, \Phi) = \langle g \circ cT, \Phi\rangle = (2\pi)^{-n}\langle F^{-1}[(Fg) * c(FT)], \Phi\rangle$$
$$= (2\pi)^{-n}\langle (FG) * c(FT),\ F^{-1}\Phi\rangle$$
$$= (2\pi)^{-n}\langle (Fg) * (E_\eta(c)FT),\ F^{-1}\Phi\rangle$$
$$= (2\pi)^{-n}\langle (Fg)_\eta, \langle (E_\eta(c)FT)_\xi, F^{-1}\Phi(\eta+\xi)\rangle\rangle$$

by Lemma 3.1.

Let $\Psi(\eta, c) = \langle (E_\eta(c)FT)_\xi, F^{-1}\Phi(\eta+\xi)\rangle$. Consider Ψ as a function of η and c a parameter, Ψ is differentiable in c in a neighborhood of 0, and

$$\frac{\partial G(c, \Phi)}{\partial c} = (2\pi)^{-n}\left\langle (Fg)_\eta, \frac{\partial}{\partial c}\Psi\right\rangle$$

(see Ref. 3, Chap. 3, Lemma 3.2). Since

$$\Psi(\eta, c) = \langle (E_\eta(c)FT)_\xi, F^{-1}\Phi(\eta+\xi)\rangle$$
$$= \langle (FT)_\xi, E_\eta(c)F^{-1}\Phi(\eta+\xi)\rangle,$$

we have

$$\frac{\partial \Psi}{\partial c} = \left\langle (FT)_\xi, \frac{\partial}{\partial c}E_\eta(c)F^{-1}\Phi(\eta+\xi)\right\rangle$$

using the same lemma as before. Therefore,

$$\frac{\partial G(c, \Phi)}{\partial c} = (2\pi)^{-n}\langle (Fg)_\eta, \langle (FT)_\xi, (\tfrac{1}{2}i)(\eta \times \xi)E_\eta(c)(F^{-1}\Phi)(\eta+\xi)\rangle\rangle.$$

The same argument as above can be used to show that

$$\frac{\partial^k G(c, \Phi)}{\partial c^k}$$
$$= (2\pi)^{-n}\langle (Fg)_\eta, \langle (FT)_\xi, (\tfrac{1}{2}i)^k(\eta \times \xi)^k E_\eta(c)F^{-1}\Phi(\eta+\xi)\rangle\rangle. \quad |||$$

Lemma 4.2:

$$\frac{\partial^k G(0, \Phi)}{\partial c^k} \Big/ k! = \langle b_k[g, T], \Phi\rangle$$

We omit the details of the proof here since the arguments given in Lemma 5.3 of Ref. 1 apply almost verbatim to the present case.

Proof of Theorem 4.1: As a function of c, $G(c, \Phi)$ is infinitely differentiable for each $\Phi \in S(R^{2n})$. Using Taylor's theorem with a remainder and expanding $G(c, \Phi)$ to N terms about 0 in powers of $c > 0$, we have

$$G(c, \Phi) = \sum_{k=0}^{N} \frac{G^{(k)}(0, \Phi)}{k!} c^N + R_N(c, \Phi, \tau(\Phi))$$

where

$$R_N(c, \Phi, \tau(\Phi)) = \frac{G^{(N+1)}(\tau(\Phi), \Phi)}{(N+1)!}\ c^{N+1},$$

$\tau(\Phi)$ is a real number depending on Φ, and $0 < \tau(\Phi) < c$. Since

$$\frac{G^{(k)}(0, \Phi)}{k!} = \langle b_k[g, T], \Phi\rangle \text{ by Lemma 4.2,}$$

we have to show that

$c^{-N}|R_N(c,\Phi,\tau(\Phi))| \to 0$ as $c \to 0$ to complete the proof.

From Lemma 4.1 we see that

$$R_N(c,\Phi,\tau(\Phi)) = (2\pi)^{-n}(\tfrac{1}{2}i)^{N+1} \times \sum (1/\alpha!)\langle(\eta^\alpha Fg) * \tau(\Phi)(\eta')^\alpha FT, F^{-1}\Phi\rangle c^{N+1};$$

therefore, if one can show

$$|(2\pi)^{-n}(\tfrac{1}{2}i)^{N+1}\sum(1/\alpha!)\langle(\eta^\alpha Fg) * \tau(\Phi)(\eta')^\alpha FT,\ F^{-1}\Phi\rangle|$$
$$\leq B(\Phi) < \infty$$

for some constant $B(\Phi) \geq 0$, then we have

$$c^{-N}|R_N(c,\Phi,\tau(\Phi))| \to 0 \quad \text{as } c \to 0.$$

Using the fact that every tempered distribution is the derivative of a continuous function slowly increasing at ∞, we can write

$$\eta^\alpha Fg = D_\xi^\beta[(1+|\xi|^2)^{k/2} g_\alpha(\xi)]$$

and

$$(\eta')^\alpha FT = D_\xi^\gamma[(1+|\xi|^2)^{t/2} f_\alpha(\xi)]$$

for some multi-indices β, γ and numbers k, $t > 0$, where g_α and f_α are bounded continuous functions on R^{2n}.

Let $\Psi = F^{-1}\Phi \in S(R^{2n})$. Thus

$$\begin{aligned}
&\langle \eta^\alpha Fg * E_\eta(\tau)(\eta')^\alpha FT, \Psi\rangle \\
&\quad = \langle (\eta^\alpha Fg)(\eta),\ \langle (E_\eta(\tau)(\eta')^\alpha FT)(\xi),\ \Psi(\xi+\eta)\rangle\rangle \\
&\quad = \langle D_\eta^\beta[(1+|\eta|^2)^{k/2} g_\alpha(\eta)], \\
&\qquad \times \langle E_\eta(\tau) D_\xi^\gamma[(1+|\xi|^2)^{t/2} f_\alpha(\xi)],\ \Psi(\xi+\eta)\rangle\rangle.
\end{aligned}$$

We denote

$$\begin{aligned}
X_\alpha(\eta) &= \langle E_\eta(\tau) D_\xi^\gamma[(1+|\xi|^2)^{t/2} f_\alpha(\xi),\ \Psi(\xi+\eta)\rangle \\
&= \langle f_\alpha(\xi),\ (1+|\xi|^2)^{t/2} D_\xi^\gamma[E_\eta(\tau)\Psi(\xi+\eta)]\rangle \\
&= \int f_\alpha(\xi)(1+|\xi|^2)^{t/2} D_\xi^\gamma[E_\eta(\tau)\Psi(\xi+\eta)]\,d\xi \\
&= \sum_{q\leq\gamma} \binom{\gamma}{q} P_q(\tau,\eta) \\
&\quad \times \int f_\alpha(\xi)(1+|\xi|^2)^{t/2} E_\eta(\tau) D_\xi^{\gamma-q}\Psi(\xi+\eta)\,d\xi,
\end{aligned}$$

where $P_q(\tau,\eta)$ is a polynomial in τ and η of degree no more than $|\gamma|$. The last equation is obtained by applying the Leibnitz rule to $D_\xi^\gamma[E_\eta(\tau)\Psi(\xi+\eta)]$ and that $D_\xi^q E_\eta(\tau) = P_q(\tau,\eta)E_\eta(\tau)$. For each τ, η, it is clear that X_α is in $S(R^{2n})$.

Likewise,

$$\begin{aligned}
&\langle D_\eta^\beta[(1+|\eta|^2)^{k/2} g_\alpha(\eta)],\ X_\alpha(\eta)\rangle \\
&\quad = \langle g_\alpha(\eta),\ (1+|\eta|^2)^{k/2} D_\eta^\beta X_\alpha(\eta)\rangle \\
&\quad = \sum_{q\leq\gamma}\sum_{p\leq\beta} \binom{\gamma}{q}\binom{\beta}{p} \langle g_\alpha(\eta),\ (1+|\eta|^2)^{k/2} P_{q\beta}(\tau,\eta) \\
&\qquad \times \langle f_\alpha(\xi),\ (1+|\xi|^2)^{t/2} D_\eta^{\beta-p}[E_\eta(\tau) D_\xi^{\gamma-q}\Psi(\xi+\eta)]\rangle\rangle \\
&\quad = \sum_{q\leq\gamma}\sum_{p\leq\beta}\sum_{u\leq\beta-p} \binom{\gamma}{q}\binom{\beta}{p}\binom{\beta-p}{u} \\
&\qquad \times \int\int g_\alpha(\eta)(1+|\eta|^2)^{k/2} P_{q\beta}(\tau,\eta) f_\alpha(\xi)(1+|\xi|^2)^{t/2} \\
&\qquad \times Q_u(\tau,\xi) D_\eta^{\beta-p-u} D_\xi^{\gamma-q}\Psi(\xi+\eta)\,d\xi\,d\eta.
\end{aligned}$$

In the above, $P_{q\beta}(\tau,\eta) = D_\eta^\beta P_q(\tau,\eta)$ and $|Q_u(\tau,\xi)| = |D_\eta^{\beta-p}E_\eta(\tau)|$ are polynomials in τ, η and τ, ξ respectively, both of degree no more than $|\beta|$.

Let

$$A(g_\alpha) = \sup_\eta |g_\alpha(\eta)| \quad \text{and} \quad A(f_\alpha) = \sup_\eta |f_\alpha(\eta)|$$

and

$$C_{q\beta}(\eta) = \sup_{0<\tau\leq 1} |P_{q\beta}(\tau,\eta)|,$$

$$C_u(\xi) = \sup_{0<\tau\leq 1} |Q_u(\tau,\xi)|.$$

$A(g_\alpha)$, $A(f_\alpha)$ are constants for each α and $C_{q\beta}(\eta)$, $D_u(\xi)$ are polynomials in η and ξ, respectively.

In these notions, we have

$$\begin{aligned}
&|\langle (\eta^\alpha Fg) * (E_\eta(\tau)(\eta')^\alpha FT),\ \Psi\rangle| \\
&\quad \leq \sum_{q\leq\gamma}\sum_{p\leq\beta}\sum_{u\leq\beta-p} \binom{\gamma}{p}\binom{\beta}{p}\binom{\beta-p}{u} A(g_\alpha)A(f_\alpha) \\
&\qquad \times \int\int (1+|\eta|^2)^{k/2}\,|C_{q\beta}(\eta)|(1+|\xi|^2)^{t/2}\,|C_u(\xi)| \\
&\qquad \times |D_\eta^{\beta-p-u} D_\xi^{\gamma-q}\Psi(\xi+\eta)|\,d\xi\,d\eta = B_\alpha(\Phi) < \infty
\end{aligned}$$

since $\Phi = F\Psi \in S(R^{2n})$. Finally, we set

$$B(\Phi) = (2\pi)^{-n}(\tfrac{1}{2})^{N+1}\sum(1/\alpha!)B_\alpha(\Phi) < \infty$$

to complete the proof. |||

*Work supported in part by a SUNY Research Foundation grant.

[1] K. C. Liu, J. Math. Phys. **16**, 724 (1975).
[2] L. Schwartz, *Théorie des distributions, I, II* (Hermann, Paris, 1957), 2nd ed.
[3] J. Barros-Neto, *An Introduction to the Theory of Distributions* (Dekker, New York, 1973).

Master equations in quantum stochastic processes

Etang Chen*

Department of Mathematics and Institute for Fundamental Studies, University of Rochester, Rochester, New York 14627
(Received 23 July 1974; revised manuscript received 11 November 1974)

Pauli master equation is derived rigorously in the framework of quantum stochastic processes.

In a series of papers[1,2] Davies has developed a theory of quantum stochastic processes to describe certain quantum measurements, where observations of the arrival of quanta at a counter are made within an interval of time. For example, optical experiments for counting photons set up in quantum optics, where measurements extend over a period of time and involve correlation functions. It is a quantum mechanical theory of stochastic processes. In this theory, the time evolution of systems is represented by a one-parameter strongly continuous semigroup of positive endomorphisms of the trace class operators, while in the classical theory, by positive real line. It has been shown that this theory corresponds most closely to the Markov jump processes in the classical theory of stochastic processes.

We shall investigate in the present note, the time evolution of quantum stochastic process in terms of (occupation) probabilities, viz. the Pauli master equation, which describes the evolution of irreversible processes in quantum systems.

Firstly, we shall derive formally, but rigorously, Pauli master equation for any "phase cell"[3] in the sample space of a quantum stochastic process. Then, we show two special cases of mater equations: (i) the phase cell is only the zero event, and (ii) the phase cell is equal to the whole sample space. It turns out that in case (ii) the "coarse-graining" is too "coarse," so that the master equation goes to zero; viz. the whole system is in a state of equilibrium. On the contrary, there is a nontrivial master equation in the case (i). Finally, we give three examples where the master equations for case (i) are shown explicitly.

A brief outline of the general formulation of quantum stochastic process is presented here. For more detailed information we refer the reader to Davies original papers.[1,2]

Each *event* is represented by a point (x, t) in $X\times(0,\infty)$, where t is the instant at which the event occurs at $x\in X$, here X is a separable locally compact Hausdorff space representing the set of all possible values of some observables. A *sample point* is a sequence of events $\{(x_i, t_i); i=1,2,\cdots\}$ such that $0<t_1<t_2<\cdots$, and either the sequence terminates or $t_n\to\infty$ as $n\to\infty$. The *sample space* X_∞ is defined as the set of all sample points, it is a Borel space. For each time $t>0$, we define the *sample space* X_t as the set of all finite sequences $\{(x_i, t_i); i=1,2,\ldots,n\}$ such that $0<t_1<t_2<\cdots<t_n\leq t$. X_t is also a Borel space.

For $0<s<t\leq\infty$, there is a Borel map π of X_t onto X_s in which each sequence $\{(x_i, t_i); i=1,2,\ldots,n\}$ is taken to the subsequence formed by dropping all the events occurring after time s. Given any $s, t>0$, there is a one-one Borel isomorphism λ from $X_s\times X_t$ onto X_{s+t} defined by

$$\lambda\{(x_i, s_i)_{i=1}^m, (y_j, t_j)_{j=1}^n\}$$
$$=(y_1, t_1)\cdots(y_n, t_n), (x_1, t+s_1)\cdots(x_m, t+s_m).$$

Let $V=T_s(H)$ be the ordered Banach space (under the trace norm) of all self-adjoint trace class operators on a Hilbert space H, and $\tau=\mathrm{tr}$, the normalized trace on $T_s(H)$. (V, τ) is the state space,[2] and $L(V)$ is the vector space of all bounded operators on V, endowed with the strong-operator topology.

A *quantum stochastic process* on X, V is a family of positive σ-additive measures $\{\mathcal{E}_t; t\geq 0\}$ on X_t with values in $L(V)$ such that

(i) $\tau[\mathcal{E}_t(X_t, \rho)]=\tau[\rho]$ for all $\rho\in V$;

(ii) $\mathcal{E}_0(X_0, \rho)=\rho$ for all $\rho\in V$, where $X_0=\{(x_t, 0)$; all $x_t\in X\}$;

(iii) for each $\rho\in V$, $t\to\mathcal{E}_t(X_t, \rho)$ is continuous for all $t\geq 0$;

(iv) for all $\rho\in V$ and $s, t>0$

$$\mathcal{E}_t(\Delta, \mathcal{E}_s(\Delta', \rho))=\mathcal{E}_t\circ\mathcal{E}_s(\Delta\times\Delta', \rho)$$
$$=\mathcal{E}_{t+s}(\lambda(\Delta\times\Delta'), \rho)$$

for Borel subsets $\Delta\subseteq X_t$ and $\Delta'\subseteq X_s$. (1)

The last condition implies that the time evolution is homogeneous. It is a generalization of the *Chapman–Kolmogorov equation*.

The physical implications of this setting are following: $\mathcal{E}_t$ is considered as a family of filters accepting a state $\rho\in V$ at time zero and emitting an output state $\mathcal{E}_t(\Delta, \rho)$ conditional on the set $\Delta\subseteq X_t$ at the latter time t. $\tau[\mathcal{E}_t(\Delta, \rho)]$, in fact $\tau[\mathcal{E}_t(\Delta, \rho)]/\tau[\rho]$ with $\tau[\rho]=1$, can be interpreted as the probability that the physical quantity can be observed in the state ρ and it takes values in Δ up to time t. In fact, $\tau[\mathcal{E}_t(\Delta, \rho)]$ is the "occupation" probability of the state ρ. Hence, Pauli master equation in quantum stochastic processes is an equation in terms of $\tau[\mathcal{E}_t(\Delta, \rho)]$.

Before studying the master equation, let us introduce some Borel subsets in the sample space X_t.

Let $\{\Delta_i\}_{i=1}^m$ be Borel subsets of X. For $0<s_1<t_1<\cdots<s_m<t_m\leq t$, let $\Delta\subseteq X_t$ be the Borel subset defined

by

$$\Delta=\{(x_i,r_i)_{i=1}^m; s_i<r_i\leq t_i \text{ and } x_i\in\Delta_i\}.$$

Then, the sets Δ defined in this form constitute a semiring, which generates the σ-field of all Borel subsets of X_t. Without loss of generality, we shall consider Borel subset of X_t only in this form. We note that Δ depends on t; and, obviously, $r_i\downarrow 0$ as $t\downarrow 0$. Now, let

$$\Delta_0=\{(x_i,0)_{i=1}^m; x_i\in\Delta_i\}. \tag{2}$$

We assume that "Δ approaches to Δ_0" as $t\downarrow 0$, denoted by $\Delta\downarrow\Delta_0$ as $t\downarrow 0$. In particular, for $\Delta=X_t$,

$$X_t\downarrow X_0 \text{ as } t\downarrow 0,$$

where $X_0=\{(x_i,0);$ all $x_i\in X\}$; and for $\Delta=z$, the zero event,

$$z\downarrow z \text{ as } t\downarrow 0.$$

We are now ready to derive master equation. For any Borel sets $\Delta\subseteq X_t$ and $\Delta'\subseteq X_s$,

$$\begin{aligned}\frac{d}{dt}\tau[\mathcal{E}_t(\Delta,\rho)]&=\lim_{s\downarrow 0}s^{-1}\tau[\mathcal{E}_{t+s}(\lambda(\Delta\times\Delta'),\rho)-\mathcal{E}_t(\Delta,\rho)]\\&=\lim_{s\downarrow 0}s^{-1}\tau[\mathcal{E}_t(\Delta,\mathcal{E}_s(\Delta',\rho))-\mathcal{E}_t(\Delta,\rho)]\\&=\lim_{s\downarrow 0}s^{-1}\tau[\mathcal{E}_t(\Delta,\mathcal{E}_s(\Delta',\rho)-\rho)],\end{aligned}$$

where the second equality is due to Eq. (1). This limit can exist if $\Delta'\downarrow\Delta_0'$ and the limit of $s^{-1}[\mathcal{E}_s(\Delta',\rho)-\rho]$ exists whenever $s\downarrow 0$. Let this limit be $A(\Delta_0')\rho$, i. e.,

$$\lim_{s\downarrow 0}s^{-1}[\mathcal{E}_s(\Delta',\rho)-\rho]=A(\Delta_0')\rho \tag{3}$$

where the limit is taken in the strong-operator topology on $L(V)$, and Δ_0' is given by Eq. (2). $A(\Delta_0')$ is a linear operator with domain

$$\mathcal{D}(A(\Delta_0'))=\{\rho\in V;\lim_{s\downarrow 0}s^{-1}[\mathcal{E}_s(\Delta',\rho)-\rho] \text{ exists in } V\}.$$

We note that $\mathcal{D}(A(\Delta_0'))$ is nonempty, it contains at least zero element. Therefore, Pauli master equation is

$$\frac{d}{dt}\tau[\mathcal{E}_t(\Delta,\rho)]=\tau[\mathcal{E}_t(\Delta,A(\Delta_0')\rho]. \tag{4}$$

This general form of master equation deserves some remarks:

First, we note that the derivation of Eq. (4) depends mainly on the assumptions of Eq. (1) and $A(\Delta_0')$ in Eq. (3). Equation (1), a generalization of Chapman–Kolmogorov equation, has the nature similar to the assumption of randomness, it is a general form of various hypotheses known as Stossansatz, molecular chaos hypothesis or random approximation. Nevertheless, the assumption of randomness (1), as we have seen above, is not sufficient to obtain Pauli master equation (4). The main reason is that Eq. (4) is Markovian type, but quantum stochastic process, in general, is not. The existence of $A(\Delta_0')$ in (3) is an additional condition which ensures a quantum stochastic process to be Markovian. Indeed, $A(\Delta_0')$, under a certain assumptions, is the infinetesimal generator of a semigroup on V (see the Appendix).

In general, the conditions given in the Appendix are not fulfilled for an arbitrary Borel subset Δ in a sample space X_t. In particular, the assumption (A7) holds only for Δ, which is either the zero event z, or the whole sample space X_t. Therefore, the possible master equations in quantum stochastic processes are only these two cases, which are derived precisely as follows:

Case (i): $\Delta=z$

We have $\mathcal{E}_t(z,\rho)=S_t(\rho)$, where $S_t: V\to V$ is a strongly continuous one-parameter semigroup on V.[1] S_t takes a pure state ρ to another pure state $S_t(\rho)$ for all $t\geq 0$.

Hence, there is a strongly continuous semigroup B_t on $\mathcal{H}$ such that $S_t(\rho)=B_t\rho B_t^*$.[1] For a pure state ρ of the positive cone V^* of V, $\rho=\xi\otimes\bar{\xi}$ for $\xi\in\mathcal{H}$.[4] Since $z\downarrow z$ as $t\downarrow 0$, thus from Eq. (3)

$$\begin{aligned}A(z)\rho&=\lim_{t\downarrow 0}t^{-1}[S_t(\rho)-\rho]\\&=\lim_{t\downarrow 0}t^{-1}[B_t(\xi\otimes\bar{\xi})B_t^*-(\xi\otimes\bar{\xi})]\\&=\lim_{t\downarrow 0}t^{-1}[(B_t\xi)\otimes\overline{(B_t\xi)}-(\xi\otimes\bar{\xi})]\\&=\lim_{t\downarrow 0}t^{-1}[(B_t\xi-\xi)\otimes\overline{(B_t\xi)}+\xi\otimes\overline{(B_t\xi-\xi)}].\end{aligned}$$

Let $Z\xi=\lim_{t\downarrow 0}t^{-1}[B_t\xi-\xi]$ for $\xi\in\mathcal{D}(Z)$, the domain of Z; i. e., Z is the infinitiesmal generator of the strongly continuous semigroup B_t on $\mathcal{H}$. We note that $\lim_{t\downarrow 0}B_t\xi=B_0\xi=\xi$ for all $\xi\in\mathcal{H}$. Then,

$$A(z)\rho=(Z\xi)\otimes\bar{\xi}+\xi\otimes\overline{(Z\xi)} \quad \text{for } \xi\in\mathcal{D}(Z).$$

Therefore,

$$\begin{aligned}\tau[\mathcal{E}_t(z,A_z(\rho))]&=\tau[S_t(A_z(\rho))]\\&=\tau[B_t\{(Z\xi)\otimes\bar{\xi}+\xi\otimes\overline{(Z\xi)}\}B_t^*]\\&=\tau[(B_tZ\xi)\otimes\overline{(B_t\xi)}+(B_t\xi)\otimes\overline{(B_tZ\xi)}]\end{aligned}$$

for all $\xi\in\mathcal{D}(Z)$. We use a fact: $B_tZ=ZB_t$ for all $\xi\in\mathcal{D}(Z)$,[5] then the Pauli master equation for zero event is

$$\begin{aligned}\frac{d}{dt}\tau[\mathcal{E}_t(z,\rho)]&=\tau[\mathcal{E}_t(z,A(z)\rho)]\\&=\tau[(ZB_t\xi)\otimes\overline{(B_t\xi)}+(B_t\xi)\otimes\overline{(ZB_t\xi)}]\\&=\langle ZB_t\xi,B_t\xi\rangle+\langle B_t\xi,ZB_t\xi\rangle\\&=2\,\mathrm{Re}\langle ZB_t\xi,B_t\xi\rangle\end{aligned} \tag{5}$$

for all $\xi\in\mathcal{D}(Z)$.

Moreover, the *total interaction rate* R is defined as a unique positive, self-adjoint bounded operator on $\mathcal{H}$ such that

$$\tau[\rho R]=\lim_{t\to\infty}\tau[\mathcal{J}_t(X,\rho)]=\tau[\mathcal{J}(X,\rho)] \tag{6}$$

where $\mathcal{J}$, called a bounded stochastic kernel on X, V, is a bounded positive σ-additive measure on the σ-field of Borel sets in X with values in $L(V)$; and[1]

$$\mathcal{J}_t(X,\rho)\equiv t^{-1}\mathcal{E}_t(A_t^1,\rho)$$

with $A_t^1=\{$all sample points in X_t containing exactly 1 events$\}$. Then, for $\xi\in\mathcal{D}(Z)$,

$$2\,\mathrm{Re}\langle Z\xi, \xi\rangle = -\langle R\xi, \xi\rangle \tag{7}$$

from a result of Davies.[1] Substituting (7) into (5), we have

$$\begin{aligned}\frac{d}{dt}\tau[\mathcal{E}_t(z,\rho)] &= -\langle RB_t\xi, B_t\xi\rangle \\ &= -\tau[R(B_t\xi)\otimes\overline{(B_t\xi)}] \\ &= -\tau[RS_t(\rho)],\end{aligned} \tag{8}$$

for $\rho = \xi\otimes\bar{\xi}$ with $\xi\in\mathcal{D}(Z)$. Let

$$K = \left\{\sum_{r=1}^{n}\alpha_r\xi_r\otimes\bar{\xi}_r;\ \xi_r\in\mathcal{D}(Z),\ n=1,2,\cdots\right\},$$

Then $K\subseteq V$. As $\xi\in\mathcal{D}(Z)$ implies $B_t\xi\in\mathcal{D}(Z)$ for all $t\geq 0$. Hence, Eq. (8) holds for all $\rho\in K$. However, K is dense in V, it is also true for all $\rho\in V$. Therefore, Eq. (8) is Pauli master equation for zero event z.

Case (ii): $\Delta = X_t$

We have $\mathcal{E}_t(X_t,\rho) = T_t(\rho)$, where $T_t: V\to V$ is a strongly continuous one-parameter semigroup on V.[1] From the definition of $\mathcal{E}_t$, $\tau[T_t(\rho)] = \tau[\rho]$ for all $\rho\in V$; therefore,

$$\frac{d}{dt}\tau[\mathcal{E}_t(X_t,\rho)] = 0. \tag{9}$$

This equation can also be obtained by our Eq. (4). In fact, we let $\Delta = \Delta' = X_t = X_s$, and note that $X_t \downarrow X_0$ as $t\downarrow 0$. Hence, from Eq. (3),

$$A(X_0)\rho = \lim_{t\downarrow 0} t^{-1}[\mathcal{E}_t(X_t,\rho) - \rho] = \lim_{t\downarrow 0} t^{-1}[T_t(\rho) - \rho]$$

is the infinitesmal generator of the semigroup T_t. Now, by the following identity[1]

$$t^{-1}[T_t(\rho) - \rho] = t^{-1}[S_t(\rho) - \rho] + \mathcal{J}_t(X,\rho) + o(t),$$

we have

$$A(X_0)\rho = A(z)\rho + \mathcal{J}(X,\rho)$$

where we have used Eq. (3) for $\Delta' = z$, and a lemma in Ref. 1 that $\mathcal{J}_t(X,\rho)$ converges in norm to $\mathcal{J}(X,\rho)$ for all $\rho\in V$ as $t\to 0$. Thus,

$$\tau[A(X_0)\rho] = \tau[A(z)\rho] + \tau[\mathcal{J}(X,\rho)].$$

Similar to case (i), let $\rho = \xi\otimes\bar{\xi}$ for $\xi\in\mathcal{D}_Z$, then

$$\begin{aligned}\tau[A(z)\rho] &= \tau[(Z\xi)\otimes\bar{\xi} + \xi\otimes\overline{(Z\xi)}] \\ &= \langle Z\xi, \xi\rangle + \langle\xi, Z\xi\rangle = 2\,\mathrm{Re}\langle Z\xi, \xi\rangle.\end{aligned}$$

On the other hand,

$$\begin{aligned}\tau[\mathcal{J}(X,\rho)] &= \tau[\rho R] \\ &= \tau[(\xi\otimes\bar{\xi})R] = \langle\xi, R\xi\rangle\end{aligned} \tag{6'}$$

for $\rho = \xi\otimes\bar{\xi}$. Therefore, from Eq. (7),

$$\tau[A(X_0)\rho] = 2\,\mathrm{Re}\langle Z\xi, \xi\rangle + \langle R\xi, \xi\rangle = 0$$

for $\rho = \xi\otimes\bar{\xi}$ with $\xi\in\mathcal{D}(Z)$. Again, by a similar argument in case (i), we have $\tau[A(X_0)\rho] = 0$ for all $\rho\in V$. Hence, Eq. (4) is now

$$\begin{aligned}\frac{d}{dt}\tau[\mathcal{E}_t(X_t,\rho)] &= \tau[\mathcal{E}_t(X_t, A(X_0)\rho)] \\ &= \tau[A(X_0)\rho] = 0\end{aligned}$$

which is Eq. (9).

In fact, $\{T_t; t\geq 0\}$ is a *dynamical semigroup* in the sense of Kossakowski,[6] who has studied extensively the existence of $A(X_t)$. However, Pauli master equation (9) is trivial; in other words, the system is in a state of equilibrium. The main reason for this situation: This case $\Delta = X_t$ is too "coarse" for the "coarse-graining" in sample space X_t.

From the viewpoint of master equations, case (i) is more interesting than case (ii). We give three examples to show different types of master equations for case (i).

Example 1: Let $X = R$, and V = the ordered Banach space of all bounded signed Borel measures on X. Define τ on V by $\tau(\mu) = \mu(X)$, then τ is a strictly positive linear functional on V, and (V,τ) is a state space.[2] The set $\{\mu\in V^*;\ \tau(\mu) = 1\}$ of normalized states on V is exactly the set of probability measures on $\mathbb{R}$. We consider a Markov jump process with linear increments.[7] Suppose that between jumps the random variables vary linearly at a rate c, and the interaction rate is given by a constant $\alpha > 0$. Finally, we assume that the jumps are described by a stochastic kernel K.

A physical example of Markov jump process is particles traveling at uniform speed through homogeneous matter occasionally scoring a collision. Each collision produces a change of energy regulated by K. The transition probabilities for the energy is a generalization of the compound Poisson distribution,[7] if the number of collision is a Poisson process. This is the case under the hypothesis concerning homogeneity of space and lack of memory. Furthermore, if it is assumed that between collisions the energy is dissipated at a constant rate due to adsorption or friction, etc., then it is a Markov jump process with linear increments.

A quantum stochastic process has been constructed for this jump process,[1] hence the master equation can be derived by Eq. (8). Indeed, S_t is defined by

$$(S_t\mu)(\Delta) = \exp(-\alpha t)\mu(\Delta - ct)$$

for $\Delta\subseteq X$ and $\mu\in V$. $\mathcal{J}$ is defined on X, V by

$$[\mathcal{J}(\Delta,\mu)](\Delta') = \alpha\int_{\Delta} K(x,\Delta')\mu(dx).$$

It can be shown that S_t is a one-parameter strongly continuous semigroup on V taking pure states to pure states; and $\mathcal{J}$ is a bounded stochastic kernel. Moreover,

$$\tau[\mathcal{J}(X,\mu)] = [\mathcal{J}(X,\mu)](X) = \alpha\int_X K(x,X)\mu\,d(x) = \alpha\tau[\mu]$$

for all $\mu\in V$. Hence, the total interaction rate $R = \alpha$. Therefore, by Eq. (8) the master equation is

$$\frac{d}{dt}\tau[\mathcal{E}_t(z,\mu)] = -\alpha\tau[S_t(\mu)] = -\alpha\exp(-\alpha t)\tau[\mu].$$

If μ is a probability measure, then rhs of the above equation is $-\alpha\exp(-\alpha t)$.

Example 2: We consider the quantum stochastic process corresponding to a finite number of detectors set up to measure the arrival of particles in a beam of fermions.[1] Let H be the antisymmetric Fock space on $H_1 = L^2(\mathbb{R}^3)\otimes\mathbb{C}^{2n+1}$ so that

$$H = \mathbb{C}\oplus H_1\oplus(H_1\otimes H_1)_a\oplus\cdots.$$

Let $x_1, \ldots, x_n \in \mathbb{R}^3$ be the positions of detectors $D_1, \ldots, D_n$, respectively. And, let $f_1, \ldots, f_n \in \mathcal{H}_1$ be concentrated in small neighborhoods of $x_1, \ldots, x_n$, respectively. Define a mapping $\varphi: \mathcal{H}_1 \to L(\mathcal{H})$ such that it takes test functions in $\mathcal{H}_1$ to the bounded annihilation operators in $L(\mathcal{H})$ associated with the representation of canonical anticommutation relation on $\mathcal{H}$. Let annihilation operator φ_i associated with the detector D_i be $\varphi(f_i)$. Suppose that D_i is sensitive to the field averaged over a small neighborhood of x_i. $\tau[\varphi_i^* \varphi_i \rho]$ for each $\rho \in T_s(\mathcal{H})^+$ and $x_i \in \mathbb{R}^3$ gives the rate of arrival of particles at x_i per unit area and unit time. If the unperturbed Hamiltonian on $\mathcal{H}$ is $\mathcal{H}_0$. Then the infinitesimal generators $\mathcal{J}$ and Z for the corresponding quantum stochastic process are given by[1]

$$\mathcal{J}(\Delta, \rho) = \sum_{i \in \Delta} \varphi_i \rho \varphi_i^*$$

for $\Delta \subseteq \mathbb{R}^3$ and

$$Z = iH_0 - \sum_{i=1}^{n} \varphi_i^* \varphi_i,$$

respectively. Therefore, the total interaction rate $R = 2\sum_{i=1}^{n} \varphi_i^* \varphi_i$. From Eq. (8), the master equation is

$$\frac{d}{dt} \tau[\mathcal{E}_t(z, \rho)] = -2 \sum_{i=1}^{n} \tau[\varphi_i^* \varphi_i S_t(\rho)],$$

which gives the total rate of arrival of particles at detectors $D_1, \ldots, D_n$ up to time t. For $\rho = \xi \otimes \bar{\xi}$ with $\xi \in \mathcal{D}(Z)$, we obtain master equation in the following form:

$$\begin{aligned} \frac{d}{dt} \tau[\mathcal{E}_t(z, \rho)] &= -2 \sum_{i=1}^{n} \tau[\varphi_i^* \varphi_i (B_t \xi) \otimes \overline{(B_t \xi)}] \\ &= -2 \sum_{i=1}^{n} \langle \varphi_i B_t \xi, \varphi_i B_t \xi \rangle \\ &= -2 \sum_{i=1}^{n} \| \varphi_i B_t \xi \|^2. \end{aligned}$$

A similar, but more explicit, form is a model of a displaced harmonic oscillator interacting with an absorptive particle detector, which is given in the following.

Example 3. Let $X = \{e\}$ and

$$\mathcal{H} = \left\{ \{\psi_n\}_{n=0}^{\infty};\ \psi \in \mathbb{C} \text{ and } \|\psi\| = \left\{ \sum_{n=0}^{\infty} |\psi_n|^2 \right\}^{1/2} < \infty \right\}.$$

Let $\mathcal{D}_\alpha$ be the space of sequences in $\mathcal{H}$ such that the norm

$$\|\psi\|_\alpha = \sup_n \{ |\psi_n| \exp(n\alpha) \}$$

is finite. Define

$$\mathcal{D} = \bigcap_{\alpha > 0} \mathcal{D}_\alpha.$$

Let ω, μ be positive numbers, Z_0 an arbitrary complex number, and a^*, a the creation and annihilation operators on $\mathcal{H}$. Then, the quantum stochastic process can be constructed[1] on $T_s(\mathcal{H})$ by using the formal Hamiltonian

$$H = \omega a^* a - \frac{1}{\sqrt{2\omega}} (\bar{Z}_0 a + Z_0 a^*)$$

and the stochastic kernel

$$\mathcal{J}(e, \rho) = \mu a \rho a^*,$$

so that

$$Z = -iH - \tfrac{1}{2} \mu a^* a.$$

Hence, the total interaction rate $R = \mu a^* a$. Therefore, the master equation is

$$\frac{d}{dt} \tau[S_t(\rho)] = -\tau[\mu a^* a S_t(\rho)].$$

In particular, we are interested in the master equation for coherent states; i. e., for every $z \in \mathbb{C}$, $\psi(z) \in \mathcal{H}$ with $\|\psi(z)\| = 1$ such that

$$\psi_n(z) = \exp(-|z|^2/2) \frac{z^n}{\sqrt{n!}}.$$

$\psi(z)$ lies in $\mathcal{D}$ and $a\{\psi(z)\} = z\psi(z)$. States of the form $\rho = \psi(z) \otimes \overline{\psi(z)}$ are called pure coherent states. Hence, for pure coherent states,

$$\frac{d}{dt} \tau[S_t(\rho)] = -\tau[\mu a^* a (B_t \psi(z)) \otimes \overline{(B_t \psi(z))}].$$

For pure coherent states, we have,[1]

$$B_t\{\psi(z)\} = \lambda(t) \psi\{z(t)\}$$

for all $t \geq 0$, where $\lambda(t)$ and $z(t)$ are two complex-valued functions of $t \geq 0$ satisfying

$$\frac{d}{dt} \{\lambda(t) \exp[|z(t)|^2/2] \psi\{z(t)\}\} = Z\{\lambda(t) \exp[|z(t)|^2/2] \psi\{z(t)\}$$

and subject to the initial conditions $z(0) = z$, $\lambda(0) = 1$. Then,

$$\begin{aligned} \frac{d}{dt} \tau[S_t(\rho)] &= -\tau[\mu a^* a (\lambda(t) \psi\{z(t)\}) \otimes \overline{(\lambda(t) \psi\{z(t)\})}] \\ &= -\mu |\lambda(t)|^2 \| a \psi\{z(t)\} \|^2 \\ &= -\mu |\lambda(t)|^2 |z(t)|^2 \| \psi\{z(t)\} \|^2. \end{aligned}$$

In particular, at $t = 0$, the right-hand side of the above equations is $-\mu |z|^2 \|\psi(z)\|^2 = -\mu |z|^2$.

ACKNOWLEDGMENT

The author wishes to thank the referee for suggestion on an earlier version of this paper.

APPENDIX

We shall show in this appendix that $A(\Delta_0')$ in Eq. (3), under a certain assumptions, is the infinitesmal generator of a unique contraction semigroup on V. First of all, let $\Delta \subseteq X_t$, $\Delta' \subseteq X_s$, and $\rho \in V$, then

$$\begin{aligned} A(\Delta_0') \mathcal{E}_t(\Delta, \rho) &= \lim_{s \downarrow 0} s^{-1} [\mathcal{E}_s(\Delta', \mathcal{E}_t(\Delta, \rho)) - \mathcal{E}_t(\Delta, \rho)] \\ &= \lim_{s \downarrow 0} s^{-1} [\mathcal{E}_{s+t}(\lambda(\Delta' \times \Delta), \rho) - \mathcal{E}_t(\Delta, \rho)]. \end{aligned}$$

On the other hand,

$$\begin{aligned} \mathcal{E}_t(\Delta, A(\Delta_0')\rho) &= \mathcal{E}_t(\Delta, \lim_{s \downarrow 0} s^{-1} [\mathcal{E}_s(\Delta', \rho) - \rho]) \\ &= \lim_{s \downarrow 0} s^{-1} [\mathcal{E}_t(\Delta, \mathcal{E}_s(\Delta, \rho)) - \mathcal{E}_t(\Delta, \rho)] \\ &= \lim_{s \downarrow 0} s^{-1} [\mathcal{E}_{t+s}(\lambda(\Delta \times \Delta'), \rho) - \mathcal{E}_t(\Delta, \rho)]. \end{aligned}$$

Consequently,

$$A(\Delta_0')\mathcal{E}_t(\Delta,\rho)=\mathcal{E}_t(\Delta,A(\Delta_0')\rho). \tag{A1}$$

Next, without loss of generality, we may assume $\tau[\rho]=1$ in the sequel. Then, $\tau[\mathcal{E}_t(\Delta,\rho)]\leq\tau[\mathcal{E}_t(X_t,\rho)]=\tau[\rho]=1$ for $\rho\in V$ and $\Delta\subseteq X_t$, and since $\mathcal{E}_t(\Delta,\rho)\in V$, we have $\|\mathcal{E}_t(\Delta,\rho)\|\leq\|\mathcal{E}_t(\Delta,\rho)\|_1=\tau[\mathcal{E}_t(\Delta,\rho)]\leq 1$.

Hence, for $n>0$, we may define formally an operator $R_n(\Delta)$ on V:

$$R_n(\Delta)\rho=\int_0^\infty \exp(-nt)\mathcal{E}_t(\Delta,\rho)\,dt \quad \text{with } \|R_n(\Delta_0)\|\leq 1/n. \tag{A2}$$

It is noted that (A2) can be defined more rigorously by means of the infinitesmal generators of quantum stochastic process[1]; however, the present form is sufficient for our purpose. Now, for $\Delta'\subseteq X_s$,

$$\begin{aligned}
&s^{-1}[\mathcal{E}_s(\Delta',R_n(\Delta)\rho)-R_n(\Delta)\rho]\\
&\quad=s^{-1}[\mathcal{E}_s(\Delta',\textstyle\int_0^\infty \exp(-nt)\mathcal{E}_t(\Delta,\rho)\,dt)\\
&\qquad-\textstyle\int_0^\infty \exp(-nt)\mathcal{E}_t(\Delta,\rho)\,dt]\\
&\quad=s^{-1}[\textstyle\int_0^\infty \exp(-nt)\mathcal{E}_{s+t}(\lambda(\Delta'\times\Delta),\rho)\,dt\\
&\qquad-\textstyle\int_0^\infty \exp(-nt)\mathcal{E}_t(\Delta,\rho)\,dt].
\end{aligned}$$

By considering $\lambda(\Delta'\times\Delta)=\Delta''$ as a Borel subset in X_{s+t}, we may change veriable $s+t$ in the first integral; and separate the second integral into $\int_0^\infty+\int_s^\infty$:

$$\begin{aligned}
&-s^{-1}\textstyle\int_0^s \exp(-nt)\mathcal{E}_t(\Delta,\rho)\,dt+s^{-1}\int_0^\infty[\exp(-n(t-s))\mathcal{E}_t(\Delta'',\rho)\\
&\quad-\exp(-nt)\mathcal{E}_t(\Delta,\rho)]\,dt.
\end{aligned}$$

If $t\to\mathcal{E}_t(\Delta,\rho)$ is continuous, then the first term tends to $-\mathcal{E}_0(\Delta_0,\rho)$ as $s\downarrow 0$. Moreover, we note that $\Delta''\downarrow\Delta$ as $s\downarrow 0$, hence $\mathcal{E}_t(\Delta'',\rho)\to\mathcal{E}_t(\Delta,\rho)$. Therefore, the second integral can be rearranged formally as follows:

$$\begin{aligned}
&s^{-1}\int_s^\infty[\exp(ns)-1]\exp(-nt)\mathcal{E}_t(\Delta,\rho)\,dt\\
&\quad=\frac{\exp(ns)-1}{s}\left[\int_0^\infty \exp(-nt)\mathcal{E}_t(\Delta,\rho)\,dt\right.\\
&\qquad\left.-\int_0^\infty \exp(-nt)\mathcal{E}_t(\Delta,\rho)\,dt\right]\\
&\quad=\frac{\exp(ns)-1}{s}\left[R_n(\Delta)\rho-\int_0^s \exp(-nt)\mathcal{E}_t(\Delta,\rho)\,dt\right],
\end{aligned}$$

which approaches to $nR_n(\Delta)\rho$ as $s\downarrow 0$, since the second term tends to 0. Therefore, we have

$$A(\Delta_0')R_n(\Delta)\rho=-\mathcal{E}_0(\Delta_0,\rho)+nR_n(\Delta)\rho \tag{A3}$$

for $\rho\in V$.

On the other hand,

$$\begin{aligned}
R_n(\Delta)A(\Delta_0')\rho&=\textstyle\int_0^\infty \exp(-nt)\mathcal{E}_t(\Delta,A(\Delta_0')\rho)\,dt\\
&=\textstyle\int_0^\infty \exp(-nt)A(\Delta_0')\mathcal{E}_t(\Delta,\rho)\,dt \qquad [\text{by (A1)}]\\
&=A(\Delta_0')\textstyle\int_0^\infty \exp(-nt)\mathcal{E}_t(\Delta,\rho)\,dt\\
&=A(\Delta_0')R_n(\Delta)\rho\\
&=-\mathcal{E}_0(\Delta_0,\rho)+nR_n(\Delta)\rho. \qquad [\text{by (A3)}]
\end{aligned}$$

Therefore,

$$R_n(\Delta)(nI-A(\Delta_0'))\rho=\mathcal{E}_0(\Delta_0,\rho)=(nI-A(\Delta_0'))R_n(\Delta)\rho;$$

thus,

$$R_n(\Delta)=(nI-A(\Delta_0'))^{-1}. \tag{A4}$$

By the continuity of $\mathcal{E}_t(\Delta,\rho)$ in t, $\|\mathcal{E}_t(\Delta,\rho)-\mathcal{E}_0(\Delta_0,\rho)\|\to 0$ at $t\downarrow 0$. Hence, for any $\epsilon>0$, we can choose a $\delta>0$ such that

$$\|\mathcal{E}_t(\Delta,\rho)-\mathcal{E}_0(\Delta_0,\rho)\|\leq\epsilon, \tag{A5}$$

for $0\leq t\leq\delta$. We have

$$nR_n(\Delta)\rho-\mathcal{E}_0(\Delta_0,\rho)=n\textstyle\int_0^\infty \exp(-nt)[\mathcal{E}_t(\Delta,\rho)-\mathcal{E}_0(\Delta_0,\rho)]\,a$$

where we have used a fact $\int_0^\infty n\exp(-nt)\,dt=1$. Then,

$$\begin{aligned}
\|nR_n(\Delta)\rho-\mathcal{E}_0(\Delta_0,\rho)\|&\leq n\textstyle\int_0^\infty \exp(-nt)\|\mathcal{E}_t(\Delta,\rho)-\mathcal{E}_0(\Delta_0,\rho)\|\\
&=n\textstyle\int_0^\delta+n\int_\delta^\infty=I_1+I_2,
\end{aligned}$$

where $\delta>0$. By (A5), for $0\leq t\leq\delta$,

$$I_1\leq\epsilon n\textstyle\int_0^\delta \exp(-nt)\,dt\leq\epsilon n\int_0^\infty \exp(-nt)\,dt=\epsilon.$$

For a fixed $\delta>0$, by $\|\mathcal{E}_t(\Delta,\rho)-\mathcal{E}_0(\Delta_0,\rho)\|\leq 2$ for $t\geq 0$, we have

$$I_2\leq 2n\textstyle\int_\delta^\infty \exp(-nt)\,dt=2\exp(-n\delta)$$

which tends to zero as $n\to\infty$. Therefore, we have established

$$\lim_{n\to\infty}nR_n(\Delta)\rho=\mathcal{E}_0(\Delta_0,\rho) \tag{A6}$$

for $\rho\in V$. In particular, if

$$\mathcal{E}_0(\Delta_0,\rho)=\rho, \quad \rho\in V. \tag{A7}$$

Then (A6) becomes $\lim_{n\to\infty}nR_n(\Delta)=\rho$ for all $\rho\in V$. And, by (A3), $R_n(\Delta)\rho\in\mathcal{D}(A(\Delta_0'))$; therefore, $\mathcal{D}(A(\Delta_0'))$ is dense in V.

Now, let $\rho_\alpha\in\mathcal{D}(A(\Delta_0'))$, $\lim_{n\to\infty}\rho_\alpha=\rho\in V$, and $\lim_{\alpha\to\infty}A(\Delta_0')\rho_\alpha=\rho'\in V$. We define $\varphi_\alpha=(I-A(\Delta_0'))\rho_\alpha$; then $\lim_{\alpha\to\infty}\varphi_\alpha=\rho-\rho'$. And, by the continuity of $(I-A(\Delta_0'))^{-1}$,

$$\rho=\lim_{\alpha\to\infty}\rho_\alpha=\lim_{\alpha\to\infty}(I-A(\Delta_0'))^{-1}\varphi_\alpha=(I-A(\Delta_0'))^{-1}(\rho-\rho');$$

hence, $(I-A(\Delta_0'))\rho=\rho-\rho'$. This implies $A(\Delta_0')\rho=\rho'$. Moreover, if (A7) holds, then, by (A4) and (A6), $\rho\in\mathcal{D}(A(\Delta_0'))$. Consequently, $A(\Delta_0')$ is a closed operator on V.

Therefore, we have shown that if $\mathcal{E}_t(\Delta,\rho)$ is continuous in t and (A7) holds, then $A(\Delta_0')$ is a closed, densely defined operator on V, and for $n>0$, n is in the resolvent set of $A(\Delta_0')$ with $\|(nI-A(\Delta_0'))^{-1}\|\leq 1/n$. Thus, by Hill–Yosida's theorem,[5] $A(\Delta_0')$ is the infinitesimal generator of a unique contraction semigroup on V.

*Research supported by NSF Grant DID-04010-A-02.

[1]E.B. Davis, Commun. Math. Phys. **15**, 277 (1969); **19**, 83 (1970); **22**, 51 (1971).
[2]E.B. Davis and J.T. Lewis, Commun. Math. Phys. **17**, 239 (1970).
[3]N.G. van Kampen, *Fundamental Problems in Statistical Mechanics I*, edited by E.G.D. Cohen (North-Holland, Amsterdam, 1962).
[4]R. Schatten, *Norm Ideals of Completely Continuous Operators* (Springer-Verlag, Berlin, 1960).
[5]K. Yosida, *Functional Analysis* (Springer-Verlag, Berlin, 1971), 3rd edition.
[6]A. Kossakowski, Bull. Acad. Polon. Sci., Sêr. Sci. Math. Astr. Phys. **20**, 1021 (1972).
[7]W. Feller, *An Introduction to Probability Theory and its Application II* (Wiley, New York, 1966).

Semisimple graded Lie algebras*

A. Pais and V. Rittenberg

Department of Physics, The Rockefeller University, New York, New York 10021
(Received 2 May 1975)

The concept of metric is introduced for graded Lie algebras. Semisimple graded Lie algebras are defined in terms of metric conditions of nonsingularity. It is shown that for this class of algebras the metric tensor generates a quadratic Casimir operator. Also for this class, the grading representation is irreducible and its weights are related to the roots of the Lie algebra ("root-weight theorem"). The problem is solved to find all semisimple graded Lie algebras. For SU(N), $N>2$, for O(N), $N>5$, and for all exceptional groups there are none. For all other semisimple Lie algebras there is one and only one. These are explicity constructed in terms of a convenient realization of Sp($2N$) matrices. SU(2) is discussed in some detail and a new group [GSU(2)] is found which leaves a mixed c-number/q-number quadratic form invariant. We also define irreducible tensor operators for this group. SU(N), $N>2$, provides examples of nonsemisimple gradings.

1. INTRODUCTION

A graded Lie algebra is an algebra which involves both commutators and anticommutators. Such an algebra is of the following general form:

$$[Q_m, Q_n] = f_{mn}^{p} Q_p, \tag{1.1}$$

$$[Q_m, V_\alpha] = F_{m\alpha}^{\beta} V_\beta, \tag{1.2}$$

$$\{V_\alpha, V_\beta\} = A_{\alpha\beta}^{m} Q_m. \tag{1.3}$$

[ ] and { } denote commutator and anticommutator respectively. The Q_m by themselves generate a Lie algebra with structure constants f_{mn}^{p}. Note that in obvious matrix notation

$$A^k = (A^k)^t, \tag{1.4}$$

where t denotes the transposed.

Algebras of this kind have made their appearance in several contexts.[1] In physics, they were first used in connection with questions related to second quantization of fermion systems.[2] They appeared again in the dual model studies by Ramond[3] and by Neveu and Schwarz.[4] In turn, this inspired Wess and Zumino[5] to employ such an algebra in connection with what is now called supersymmetry, a subject which was subsequently elaborated further by them and by a number of other authors.[6] In many of these instances the algebra is Abelian in the sense that $f_{ij}^{k} = 0$, $F_{i\alpha}^{\beta} = 0$. For example, for (global) supersymmetries the Q's are the space–time translation operators.

It is the purpose of this paper to derive a number of results which bear on the non-Abelian case. Our original motivation for this work was to find possible alternative ways to incorporate internal symmetries into supersymmetry considerations. This did not lead us anywhere. However, during these studies we discovered that these non-Abelian structure are remarkably tight and fairly elegant. This leads us to believe that the communication of these mathematical results may be of some interest in its own right.

Along with the Jacobi identity

$$[Q_m, [Q_n, V_\alpha]] + [V_\alpha, [Q_m, Q_n]] + [Q_n, [V_\alpha, Q_m]] = 0, \tag{1.5}$$

these graded algebras contain two related identities, namely,

$$[Q_m, \{V_\alpha, V_\beta\}] + \{[V_\alpha, Q_m], V_\beta\} + \}[V_\beta, Q_m], V_\alpha\} = 0, \tag{1.6}$$

$$[V_\alpha, \{V_\beta, V_\gamma\}] + [V_\gamma, \{V_\alpha, V_\beta\}] + [V_\beta, \{V_\gamma, V_\alpha\}] = 0. \tag{1.7}$$

Note that Eqs. (1.6), (1.7) are trivial in the Abelian case. On the other hand, they are the main origin for the strongly constrained character of the non-Abelian graded algebras, as we shall see in the sequel. With the help of Eqs. (1.1)–(1.3) these two identities can be written as follows:

$$A^m F_n + F_n^t A^m = f_{np}^{m} A^p \tag{1.8}$$

(where we have used matrix notation[7]), and

$$A_{\alpha\beta}^{p} F_{p\gamma}^{\delta} + A_{\gamma\alpha}^{p} F_{p\beta}^{\delta} + A_{\beta\gamma}^{p} F_{p\alpha}^{\delta} = 0. \tag{1.9}$$

Our own curiosity in this area of problems was aroused initially by a study of SU(2), which led to results which to us were surprising. Since this example is elementary and instructive, we think it helpful to begin our expose with a detailed discussion of this case in Sec. 2.

There we note that if V_α is irreducible, there is one and only one possible choice for it, namely the two-dimensional spinor representation. We then find that the graded SU(2) algebra has a quadratic Casimir operator K_2 given by

$$K_2 = -Q_i^2 - V^t C^{-1} V. \tag{1.10}$$

Here Q_i^2 is, of course, the angular momentum. $C = i\tau_2$ is the familiar 2×2 "charge conjugation matrix." Generally, for any group, we define a Casimir operator K for a graded algebra by the conditions that

$$[K, Q_i] = [K, V_\alpha] = 0 \tag{1.11}$$

shall be nontrivially satisfied. We shall see in Sec. 2 that the existence of the quadratic K, Eq. (1.10), is important for proving the full reducibility of the representations. The latter turn out to be fully labeled by J, I, I_3, where I, I_3 are the familiar angular momentum labels, while for an irreducible representation $I = J$ or $J - \frac{1}{2}$. We prove that the Schur lemma holds and construct the Clebsch–Gordan series for this algebra. Finally we make the transition from this algebra to a corresponding *new group* which we call GSU(2). It turns out that this group is defined as the set of linear transformations which leaves invariant a specific bilinear

form [see Eq. (2.19) below]. We define tensor operators for GSU(2) and check the existence of a Wigner–Eckart theorem for an example.

Thereupon, we made forays into other Lie groups and their gradings, but soon we came to the realization that we needed some organizing principle to systematize our explorations. This led us to introduce the notion of semisimplicity for graded Lie algebras. As we shall show in Sec. 3, there exists a quite natural definition of a metric tensor for a graded Lie algebra. This tensor acts in a manifold of $D+d$ dimensions, where D is the dimension of the Lie group and d the dimension of V_α. The precise definition of this tensor $g_{\mu\nu}$, $\mu,\nu = 1,\ldots,D+d$ is given in Eqs. (3.7), (3.8) below. Along with $g_{\mu\nu}$, it is convenient to introduce also the familiar metric tensor h_{mn} of the underlying Lie algebra:

$$h_{mn} = f^p_{mq} f^q_{np}, \tag{1.12}$$

where all indices run over the range $1,\ldots,D$. Let us now consider four classes of graded Lie algebras, defined as follows:

(I) $\det|g_{\mu\nu}| \neq 0$, $\det|h_{mn}| \neq 0$. We shall refer to a member of this class as a *semisimple graded Lie algebra*. It is the main result of this paper that we have found all semisimple graded Lie algebras.

(II) $\det|g_{\mu\nu}| = 0$, $\det|h_{mn}| \neq 0$. Here the question is: Can an SSLA [8] be graded in such a way that the resulting SSGLA is not semisimple? This is indeed possible and examples will be given (Sec. 6). However, we have not found a way to classify exhaustively algebras of this kind.

(III) $\det|g_{\mu\nu}| = 0$, $\det|h_{mn}| = 0$. Here we enter the domain of non-SSLA. Examples of this type of grading can easily be given (Appendix B).

(IV) $\det|g_{\mu\nu}| \neq 0$, $\det|h_{mn}| = 0$. We have found neither an example for this category nor a general argument that such cases do not exist. Therefore, it remains an unresolved question at this time whether it would or would not be sufficient to specify an SSGLA by the condition $\det|g_{\mu\nu}| \neq 0$ only.

Section 3 is devoted to the development of a number of theorems which greatly facilitate the task of finding all SSGLA. The main findings of that section are as follows.

(a) For each SSGLA there exists a quadratic Casimir operator K_2. The relevance of this result for full reducibility of GSU(2) has already been mentioned. We conjecture that full reducibility holds for any SSGLA.

(b) For each SSGLA, V_α is an irreducible representation of the corresponding SSLA.

(c) *The C-theorem*: For each SSGLA there exists necessarily a matrix C such that

$$A^k = CF_k \quad \text{and} \quad C^t = -C. \tag{1.13}$$

The C-matrix enters in K_2. Indeed, in a coordinate system in which the SSLA metric tensor is the unit matrix, K_2 is of the form given by Eq. (1.10).

(d) *The root-weight theorem*: Using the Cartan basis, let a denote generically any of the root vectors of the SSLA. Likewise let α denote any of the weight vectors of the V_α-representation. The theorem says that for any SSGLA there should necessarily exist for each α a corresponding a such that

$$a = 2\alpha. \tag{1.14}$$

From this theorem we further derive the relation

$$c_2(A) = 2c_2(V) + 2, \tag{1.15}$$

where $c_2(A)$ is the value of the quadratic Casimir operator in the adjoint representation of the SSLA and $c_2(V)$ is the same quantity for the grading representation V_α. In Eq. (1.15) the norms of the generators are relevant. Equation (1.15) holds for the norms used by Racah. [9]

The C-theorem and the root-weight theorem are necessary conditions for semisimple grading. They are not sufficient conditions. Even so, these theorems are powerful, since they enable us to exclude vast numbers of options for V_α, for any SSLA. Indeed, with their help we show the following in Sec. 4:

(1) For any SU(N), $N \geq 3$, semisimple grading is impossible.

(2) The same is true for all exceptional Lie algebras.

(3) The same is true for all orthogonal groups $O(N)$, $N \geq 6$. For each of the cases $N = 3, 4, 5$ there is one and only one semisimple grading candidate, namely the corresponding spinor representation.

(4) For each Sp($2N$) there is again one and only one such candidate, namely the fundamental $2N$-dimensional representation. From well-known local isomorphisms, the possible SS gradability of O(3), O(4), O(5) is, of course, linked to the same property of Sp(2) and Sp(4).

The derivations in Sec. 4 were facilitated enormously by the work of Mehta and Srivastava[10] who have given for each semisimple Lie algebra: (a) the dominant weights and the root vectors, (b) the representations for which $C^t = -E$.

Having recognized that there are only a limited number of possibilities left, we have the remaining task to verify whether these cases are actually realizable. Here the answer is affirmative. The explicit constructive proofs for these instances are given in Sec. 5. For this purpose we give a very convenient representation for symplectic matrices, see Eq. (5.10). Gradability for the groups just mentioned also applies to their noncompact versions. Thus for example the Lorentz group is gradable.

In Sec. 6 we discuss an interesting family of examples where grading is possible, but not in a semisimple fashion. Namely, the cases considered have the property that the metric tensor of the graded algebra vanishes (so that a fortiori its determinant is zero). These examples are SU(N), $N \geq 3$, where grading can be implemented with V_α = the adjoint representation.

Here we note that there has been recent discussion[11] elsewhere of graded Lie algebras, including comments on symplectic and Clifford algebras and on the grading just mentioned of SU(N).

Finally, we list in Sec. 7 some major questions which deserve further study.

2. GSU(2)

A. Irreducible representations of the algebra

Let Q_m $(m=1,2,3)$ be the generators of the SU(2) algebra:

$$[Q_m, Q_n] = i\epsilon_{mnp}Q_p. \tag{2.1}$$

This algebra can be graded using a spinor V_α $(\alpha = \frac{1}{2}, -\frac{1}{2})$ as a grading representation:

$$[Q_m, V_\alpha] = \tfrac{1}{2}\tau^m_{\beta\alpha}V_\beta, \tag{2.2}$$

$$\{V_\alpha, V_\beta\} = \tfrac{1}{2}(C\tau^m)_{\alpha\beta}Q_m, \tag{2.3}$$

where

$$\tau_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \tau_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \tau_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad C = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{2.4}$$

The commutation relations (2.1), (2.2) and the anti-commutation relations (2.3) satisfy the identities (1.6), (1.7), as can easily be checked.

We define

$$Q_\pm = Q_1 \pm iQ_2. \tag{2.5}$$

With the help of the generators $Q_\pm$ and Q_3, the relation (2.3) reads

$$V^2_{1/2} = \tfrac{1}{4}Q_+, \quad V^2_{-1/2} = -\tfrac{1}{4}Q_-, \quad \{V_{1/2}, V_{-1/2}\} = -\tfrac{1}{2}Q_3. \tag{2.6}$$

One observes that the square of each odd generator (V_α) is equal to an even generator $(Q_\pm)$. As will be shown in Sec. 3, this property generalizes for any SSGLA.

We now ask for a matrix representation of the Q_m and V_α which satisfies the relations (2.1)–(2.3). For the generators Q_m we have

$$\begin{aligned} &\langle I, I_3+1 | Q_+ | I, I_3\rangle = \sqrt{(I-I_3)(I+I_3+1)}, \\ &\langle I, I_3-1 | Q_- | I, I_3\rangle = \sqrt{(I+I_3)(I-I_3+1)}, \\ &\langle I, I_3 | Q_3 | I, I_3\rangle = I_3. \end{aligned} \tag{2.7}$$

For the generators V_α we get from the Wigner–Eckart theorem

$$\begin{aligned} &\langle I+\tfrac{1}{2}, I_3+\tfrac{1}{2} | V_{1/2} | I, I_3\rangle = c(I)\sqrt{I+I_3+1}, \\ &\langle I-\tfrac{1}{2}, I_3+\tfrac{1}{2} | V_{1/2} | I, I_3\rangle = d(I)\sqrt{I-I_3}, \\ &\langle I+\tfrac{1}{2}, I_3-\tfrac{1}{2} | V_{-1/2} | I, I_3\rangle = c(I)\sqrt{I-I_3+1}, \\ &\langle I-\tfrac{1}{2}, I_3-\tfrac{1}{2} | V_{-1/2} | I, I_3\rangle = -d(I)\sqrt{I+I_3}. \end{aligned} \tag{2.8}$$

Here $c(I)$ and $d(I)$ are as yet unknown functions. Introducing the expressions (2.7) and (2.8) into (2.6), we find the following equations for them:

$$\begin{aligned} &c(I-\tfrac{1}{2})d(I) + c(I)d(I+\tfrac{1}{2}) = \tfrac{1}{4}, \\ &c(I)c(I+\tfrac{1}{2}) = 0, \\ &d(I)d(I-\tfrac{1}{2}) = 0. \end{aligned} \tag{2.9}$$

From Eqs. (2.9) we conclude that an irreducible representation of the algebra GSU(2) contains a doublet: one representation with spin $I=J$ and another one with spin $I=J-\frac{1}{2}$ and that the solution of Eqs. (2.9) is (with a convenient choice of phases)

$$\begin{aligned} &c(J) = 0, \quad c(J-\tfrac{1}{2}) = -\tfrac{1}{2}, \\ &d(J) = -\tfrac{1}{2}, \quad d(J-\tfrac{1}{2}) = 0. \end{aligned} \tag{2.10}$$

We will show later that J is related to the eigenvalues of the quadratic Casimir operator and that J labels the IR of the GSU(2) algebra. Within a given representation the states are labelled by I and I_3. We thus write these states as $|J, I, I_3\rangle$. The matrix elements of the generators V_α can then be written as

$$\begin{aligned} &\langle J, J-\tfrac{1}{2}, I_3+\tfrac{1}{2} | V_{1/2} | J, J, I_3\rangle = -\tfrac{1}{2}\sqrt{J-I_3}, \\ &\langle J, J, I_3+\tfrac{1}{2} | V_{1/2} | J, J-\tfrac{1}{2}, I_3\rangle = -\tfrac{1}{2}\sqrt{J+I_3+\tfrac{1}{2}}, \\ &\langle J, J-\tfrac{1}{2}, I_3-\tfrac{1}{2} | V_{-1/2} | J, J, I_3\rangle = \tfrac{1}{2}\sqrt{J+I_3}, \\ &\langle J, J, I_3-\tfrac{1}{2} | V_{-1/2} | J, J-\tfrac{1}{2}, I_3\rangle = -\tfrac{1}{2}\sqrt{J-I_3+\tfrac{1}{2}}. \end{aligned} \tag{2.11}$$

The matrix elements of the generators Q_m are given by Eq. (2.7) with $I=J$ and $I=J-\frac{1}{2}$, $J=0, \frac{1}{2}, 1, \cdots$.

The adjoint representation corresponds to $J=1$. For this representation we easily derive from Eq. (2.11)

$$\begin{aligned} &(V_\alpha)_{m,\beta} = \tfrac{1}{2}(C\tau^m)_{\alpha\beta}, \quad (V_\alpha)_{\beta,m} = -\tfrac{1}{2}\tau^m_{\beta\alpha}, \\ &(V_\alpha)_{\beta\gamma} = (V_\alpha)_{mn} = 0. \end{aligned} \tag{2.12}$$

The quadratic Casimir operator for the GSU(2) algebra is:

$$K_2 = \mathbf{Q}^2 + C_{\alpha\beta}V_\alpha V_\beta. \tag{2.13}$$

Indeed, using (2.1)–(2.3), one can check that

$$[K_2, Q_m] = [K_2, V_\alpha] = 0. \tag{2.14}$$

From Eqs. (2.7) and (2.11) we can derive the relation

$$\langle J, I, I_3 | K_2 | J, I, I_3\rangle = J(J+\tfrac{1}{2}). \tag{2.15}$$

Thus the eigenvalues of K_2 are related to J. Since we have proven that the IR of the GSU(2) algebra are labelled with the help of only one Casimir operator we can repeat Racah's[9] proof of the total reducibility of the representations of SU(2) to show the *total reducibility* of the representations of GSU(2). This we do next.

If the representation is reducible and has for example two irreducible constituents, the generators may be brought to the form

$$Q_m = \begin{pmatrix} Q^{(1)}_m & K_m \\ 0 & Q^{(2)}_m \end{pmatrix}, \quad V_\alpha = \begin{pmatrix} V^{(1)}_\alpha & K_\alpha \\ 0 & V^{(2)}_\alpha \end{pmatrix}. \tag{2.16}$$

Let us assume for example that the Casimir operator K_2 has different eigenvalues say $\lambda^{(1)} \neq \lambda^{(2)}$ for the two IR:

$$K_2 = \begin{pmatrix} \lambda^{(1)}\mathbb{1} & K \\ 0 & \lambda^{(2)}\mathbb{1} \end{pmatrix}, \tag{2.17}$$

where K is some matrix. The transformation

$$T = \begin{pmatrix} \mathbb{1} & K/(\lambda^{(1)} - \lambda^{(2)}) \\ 0 & \mathbb{1} \end{pmatrix} \tag{2.18}$$

diagonalizes K_2 and Q_m, V_α since they commute with K_2:

$$TK_2T^{-1} = \begin{pmatrix} \lambda^{(1)}\mathbb{1} & 0 \\ 0 & \lambda^{(2)}\mathbb{1} \end{pmatrix}.$$

Finally we show the equivalent of the *Schur lemma*: Let J be an irreducible representation of the GSU(2) algebra and Q_m, V_α the generators acting on this representation. Then any matrix A which commutes with Q_m and V_α is a multiple of the unit matrix.

Proof: We choose the basis as in Eq. (2.11). In this

basis the generators are

$$Q_m = \begin{pmatrix} Q_m^J & 0 \\ 0 & Q_m^{J-1/2} \end{pmatrix}, \quad V_\alpha = \begin{pmatrix} 0, & V_\alpha^{J,J-1/2} \\ V_\alpha^{J-1/2,J}, & 0 \end{pmatrix},$$

where Q_m^J, $(Q_m^{J-1/2})$ are the matrices corresponding to Q_m acting on the $J(J-\frac{1}{2})$ irreducible representation of the SU(2) algebra; the meaning of $V_\alpha^{J,J-1/2}$, $V_\alpha^{J-1/2,J}$ is obvious. From the condition $[A, Q_m]=0$ and the application of the Schur lemma to SU(2) we have

$$A = \begin{pmatrix} \lambda \cdot \mathbb{1} & \\ 0 & \mu \cdot \mathbb{1} \end{pmatrix},$$

where λ and μ are numbers. We now use $[A, V_\alpha]=0$ and get $\lambda=\mu$, which proves the lemma.

B. The group GSU(2)

We consider next the *group* GSU(2), defined as the set of all linear transformations which leave invariant the quadratic form

$$\mathbf{x}^2 + C_{\alpha\beta} y_\alpha y_\beta. \tag{2.19}$$

Here the x_m, $m=1,2,3$, are c-numbers, but the y_α, $\alpha=1,2$, are a pair of anticommuting variables,

$$\{y_\alpha, y_\beta\} = 0. \tag{2.20}$$

Since several novel features are involved in the construction of the group elements, we shall proceed stepwise toward the general answer.

Step 1: We collect the x_m and y_α into a five-component quantity Z_μ:

$$Z_\mu = (x_m, y_\alpha), \quad \mu = 1, \ldots, 5, \tag{2.21}$$

and consider the transformation

$$Z'_\mu = Z_\nu(T(\theta))_{\mu\nu}, \tag{2.22}$$

where

$$T(\theta) = \exp(\theta_\alpha V_\alpha). \tag{2.23}$$

In Eq. (2.23), the quantities θ_α are the "fundamental parameters," a pair of anticommuting objects:

$$\{\theta_\alpha, \theta_\beta\} = 0. \tag{2.24}$$

Moreover, we postulate that

$$\{\theta_\alpha, y_\beta\} = \{\theta_\alpha, V_\beta\} = 0. \tag{2.25}$$

The $\mu\nu$-matrix elements of $T(\theta)$ are fully specified by $(V_\alpha)_{\mu\nu}$. In turn, the latter are given by Eq. (2.12). After some algebra, one sees that Eq. (2.22) can be written in the following long-hand form:

$$\begin{aligned} \mathbf{x}' &= (1 - \tfrac{1}{8} C_{\alpha\beta}\theta_\alpha\theta_\beta)\mathbf{x} - \tfrac{1}{2}(C\boldsymbol{\tau})_{\alpha\beta}\theta_\alpha y_\beta, \\ y'_\alpha &= (1 + \tfrac{3}{16} C_{\gamma\delta}\theta_\gamma\theta_\delta) y_\alpha - \tfrac{1}{2}(\boldsymbol{\tau}\mathbf{x})_{\alpha\beta}\theta_\beta, \end{aligned} \tag{2.26}$$

and it is readily verified that Eq. (2.22), or equivalently Eq. (2.26), leaves Eq. (2.19) invariant. Observe that because of the anticommutation properties, Eqs. (2.24) and (2.25), the order of the Z and the T factor on the rhs of Eq. (2.22) is essential.

We conclude this first step by noting the connection, similar to what happens for ordinary Lie groups, between the generators V and derivatives involving the T transformation:

$$(V_\alpha)_{\mu\nu} = \frac{\partial^2 Z'_\mu}{\partial\theta_\alpha \partial Z_\nu}\bigg|_{\theta_\beta=0}. \tag{2.27}$$

Because of anticommutativities this second derivative needs further specification. The rules are: If $\{u, v\}=0$, then

$$\frac{\partial}{\partial u}(uv) = v, \quad \frac{\partial}{\partial v}(uv) = -u, \quad \frac{\partial^2}{\partial u\,\partial v} f(u,v) = \frac{\partial}{\partial u}\left(\frac{\partial}{\partial v} f(u,v)\right).$$

Step 2: From the single transformation $T(\theta)$ we derive next a continuum of transformations which leave Eq. (2.19) invariant, in the following way. Consider the set λ_α defined by

$$\lambda = D\theta, \tag{2.28}$$

where D is any arbitrary 2×2 complex valued c-number matrix. Clearly λ satisfies the same relations (2.24) and (2.25) as does θ. Note also that $\{\lambda_\alpha, \theta_\beta\}=0$. From the foregoing it is obvious that all transformations $Z'_\mu = Z_\nu(T(\lambda))_{\mu\nu}$ also leave Eq. (2.19) invariant. We next make the following comments.

(1) Since the set of D matrices occurring in Eq. (2.28) includes the null matrix, we can define a unit element:

$$T(0) = 1. \tag{2.29}$$

(2) We must ask how the product $T(\lambda')T(\lambda)$ acts. One finds

$$T(\lambda')T(\lambda) = T(\lambda'+\lambda)\exp(\tfrac{1}{4}\lambda^t C\tau^k\lambda' Q_k), \tag{2.30}$$

in which now the SU(2) generators Q_k make their appearance: The product of two T transformations is in general a T transformation times a rotation. However, if $\lambda' = \text{const}\cdot\lambda$, then $\lambda^t C\tau^k\lambda' = 0$ and in particular

$$T(-\lambda)T(\lambda) = 1, \tag{2.31}$$

so that we have a well-defined inverse of $T(\lambda)$.

Note that because of Eq. (2.28), we can write

$$\theta^t C\tau^k\lambda Q_k = \beta_k\theta_1\theta_2 Q_k, \tag{2.32}$$

where the *complex* numbers β_k are given by

$$\beta_1 = D_{12} + D_{21}, \quad \beta_2 = i(D_{12} - D_{21}), \quad \beta_3 = D_{11} - D_{22}.$$

Equation (2.32) draws our attention to the fact that the parameters associated with Q_k may in general contain two parts (both even in the θ's): one which is independent of the θ's and one which is proportional to $\theta_1\theta_2$.

Step 3: We are now ready to write down the most general transformation which leaves Eq. (2.19) invariant. It is given by

$$\begin{aligned} Z'_\mu &= Z_\nu(\exp(\lambda_\alpha V_\alpha)\cdot\exp(\gamma_m Q_m))_{\mu\nu} \\ &= Z_\nu(T(D\theta)\exp[(\alpha_m + \beta_m\theta_1\theta_2)\theta_m])_{\mu\nu}, \end{aligned} \tag{2.33}$$

where D has been described above. The α_m are pure imaginary. They are, of course, the pure SU(2) transformations. The β_m are complex.

The additional labor to show that these transformations form a group goes by routine steps. One formally expands the various exponentials and makes reductions with the help of the relations Eqs. (2.1)–(2.3). The group GSU(2) is the first example, as far as we know, of a *non-Abelian graded group* in the sense of Berezin and Kac.[2]

The quadratic form (2.19) is also left invariant by a new type of parity operation (we call it "S" parity) which corresponds to the transformation

$$x \to x, \quad y_\alpha \to -y_\alpha, \tag{2.34}$$

in which the anticommuting variables change sign. As seen from (2.33) this corresponds to an outer automorphism of the group GSU_2.

From Eq. (2.34) we derive

$$SQ_mS^{-1} = Q_m, \quad SV_\alpha S^{-1} = -V_\alpha, \tag{2.35}$$

$$S^2 = 1 \tag{2.36}$$

It is easy to show that for IR of GSU(2), the operator S has the expression

$$S = W(4W+1)(8W+1)/K_2(16K_2+1)^{1/2}, \tag{2.37}$$

where

$$W = C_{\alpha\beta}V_\alpha V_\beta$$

and

$$S|J, J, I_3\rangle = |J, J, I_3\rangle,$$
$$S|J, J-\tfrac{1}{2}, I_3\rangle = -|J, J-\tfrac{1}{2}, I_3\rangle. \tag{2.38}$$

Thus the IR of the group GSU(2) are also IR of the union of this group with the S-parity operation.

C. Tensor operators

The Clebsch–Gordan series for the product of two representations of the algebra (2.1)–(2.3) reads:

$$J \otimes J' = |J-J'|, |J-J'|+\tfrac{1}{2}, \ldots, J+J'-\tfrac{1}{2}, J+J'. \tag{2.39}$$

We do not give the general proof of Eq. (2.39), but we will illustrate presently not only how Eq. (2.39) works for the simple example

$$J \otimes \tfrac{1}{2} = |J-\tfrac{1}{2}|, J, J+\tfrac{1}{2}, \tag{2.40}$$

but also the Clebsch–Gordan coefficients for this case will be given.

We define *irreducible tensors* T^J_{I,I_3} by the following commutation and anticommutation relations:

$$[Q_m, T^J_{I,I_3}] = \sum_{I'_3} \langle J, I, I'_3|Q_m|J, I, I_3\rangle T^J_{I,I'_3},$$
$$[V_\alpha, T^J_{J,I_3}] = \langle J, J-\tfrac{1}{2}, I_3+\alpha|V_\alpha|J, J, I_3\rangle T^J_{J-1/2,I_3+\alpha}, \tag{2.41}$$
$$\{V_\alpha, T^J_{J-1/2,I_3}\} = \langle J, J, I_3+\alpha|V_\alpha|J, J-\tfrac{1}{2}, I_3\rangle T^J_{J,I_3+\alpha}.$$

We conjecture that an analog of the Wigner–Eckart theorem holds true, namely

$$\langle J'', I'', I''_3|T^{J'}_{I',I'_3}|J, I, I_3\rangle = (J''\|T^{J'}\|J) C^{J'',I',I''}_{J,I,I_3;J',I',I'_3}. \tag{2.42}$$

On the right-hand side of Eq. (2.42), the first term represents the reduced matrix element and the second one the Clebsch–Gordan coefficient. We have not given a general proof of Eq. (2.42), but we have checked its validity for the following example.

Let us take $J = \frac{1}{2}$ in (2.41). With the help of Eqs. (2.7) and (2.11) one obtains

$$[Q_+, T^{1/2}_{1/2,1/2}] = [Q_\pm, T^{1/2}_{0,0}] = [Q_3, T^{1/2}_{0,0}] = [Q_-, T^{1/2}_{1/2,-1/2}] = 0,$$
$$[Q_\pm, T^{1/2}_{1/2,\mp1/2}] = T^{1/2}_{1/2,\pm1/2}, \quad [Q_3, T^{1/2}_{1/2,\pm1/2}] = \pm\tfrac{1}{2}T^{1/2}_{1/2,\pm1/2}, \tag{2.43}$$

$$[V_{\pm1/2}, T^{1/2}_{1/2,\pm1/2}] = 0, \quad [V_{\pm1/2}, T^{1/2}_{1/2,\mp1/2}] = \mp\tfrac{1}{2}T^{1/2}_{0,0}, \tag{2.44}$$

$$\{V_{\pm1/2}, T^{1/2}_{0,0}\} = -\tfrac{1}{2}T^{1/2}_{1/2,\pm1/2}. \tag{2.45}$$

Equations (2.43)–(2.45) have been solved, and we have verified that they have the structure as in Eq. (2.42). By further imposing the orthonormality relations, we have determined the Clebsch–Gordan coefficients as follows:

$$C^{J+1/2,J+1/2,I_3\pm1/2}_{J,J,I_3;1/2,1/2,\pm1/2} = c_1\sqrt{J\pm I_3+1},$$
$$C^{J+1/2,J,I_3\pm1/2}_{J,J-1/2,I_3;1/2,1/2,\pm1/2} = c_1\sqrt{J\pm I_3+\tfrac{1}{2}}$$
$$C^{J+1/2,J,I_3}_{J,J,I_3;1/2,0,0} = c_1,$$
$$C^{J-1/2,J-1/2,I_3\pm1/2}_{J,J,I_3;1/2,1/2,\pm1/2} = \pm c_2\sqrt{J\mp I_3},$$
$$C^{J-1/2,J-1/2,I_3\pm1/2}_{J,J-1/2,I_3;1/2,1/2,\pm1/2} = \pm c_2\sqrt{J\mp I_3-\tfrac{1}{2}}, \tag{2.46}$$
$$C^{J-1/2,J-1/2,I_3}_{J,J-1/2,I_3;1/2,0,0} = c_2,$$
$$C^{J,J-1/2,I_3\pm1/2}_{J,J,I_3;1/2,1/2,\pm1/2} = \pm c_3\sqrt{J\mp I_3},$$
$$C^{J,J,I_3\pm1/2}_{J,J-1/2,I_3;1/2,1/2,\pm1/2} = c_3\sqrt{J\pm I_3+\tfrac{1}{2}},$$
$$C^{J,J,I_3}_{J,J,I_3;1/2,0,0} = -2Jc_3, \quad C^{J,J-1/2,I_3}_{J,J-1/2,I_3;1/2,0,0} = (2J+1)c_3,$$

where

$$c_1 = \pm 1/\sqrt{2J+1}, \quad c_2 = \pm 1/\sqrt{2J}, \quad c_3 = \pm 1/\sqrt{2J(2J+1)}. \tag{2.47}$$

The entire discussion of this section was based on the choice of a spinor for the grading representation V. We must finally ask if there is any other possible choice for V. By following exactly the techniques outlined above, one readily finds that any *irreducible* SU(2) representation for V other than a spinor does not work. We have not pursued in all generality the possible use of reducible representations. However, for the specific class of semisimple graded Lie algebras, V *has* to be irreducible. We turn next to the derivation of general properties of this class of algebras. Our GSU(2) example discussed in the present section will soon be seen to belong to this class.

3. A STANDARD FORM FOR SEMISIMPLE GRADED LIE ALGEBRAS

The commutation and anticommutation relations (1.1)–(1.3) and the identities Eqs. (1.5)–(1.7) can be cast in a more compact form as follows.[1,2] Let X_μ comprise the sets of generators Q_m, $(m = 1, \ldots, D)$ and V_α $(\alpha = 1, \ldots, d)$:

$$X_\mu = Q_m, V_\alpha, \quad \mu = 1, \ldots, D+d. \tag{3.1}$$

Further, define the degree $g(X_\mu)$ of a generator by

$$g(Q_m) = 0, \quad g(V_\alpha) = 1. \tag{3.2}$$

If $g = 0$ (1), the corresponding operator will be called even (odd). Define (X_μ, X_ν) by

$$(X_\mu, X_\nu) \equiv X_\mu X_\nu - (-1)^{g(X_\mu)g(X_\nu)}X_\nu X_\mu. \tag{3.3}$$

Then Eqs. (1.1)–(1.3) can be comprised as

$$(X_\mu, X_\nu) = c^\omega_{\mu\nu}X_\omega. \tag{3.4}$$

The c-symbols satisfy

$$c^\omega_{\mu\nu} = -(-1)^{g(X_\mu)g(X_\nu)}c^\omega_{\nu\mu}. \tag{3.5}$$

Equations (1.5)–(1.7) are all contained in

$$(X_\mu,(X_\nu,X_\omega))\cdot(-1)^{\varepsilon(X_\mu)\varepsilon(X_\omega)}+(X_\nu,(X_\omega,X_\mu))\cdot(-1)^{\varepsilon(X_\nu)\varepsilon(X_\mu)}$$
$$+(X_\omega,(X_\mu,X_\nu))\cdot(-1)^{\varepsilon(X_\omega)\varepsilon(X_\nu)}=0. \tag{3.6}$$

We define a metric tensor $g_{\mu\nu}$ for the graded algebra by

$$g_{\mu\nu}=c^{\sigma}_{\omega\mu}\cdot(-1)^{\varepsilon(X_\omega)}c^{\omega}_{\sigma\nu}=(-1)^{\varepsilon(X_\mu)\varepsilon(X_\nu)}g_{\nu\mu}. \tag{3.7}$$

Equivalently [recall that h_{mn} was defined in Eq. (1.12)],

$$\begin{aligned} g_{mn}&=h_{mn}-F^{\beta}_{m\alpha}F^{\alpha}_{n\beta}=g_{nm},\\ g_{\alpha\beta}&=F^{\gamma}_{m\alpha}A^{m}_{\beta\gamma}-F^{\gamma}_{m\beta}A^{m}_{\alpha\gamma}=-g_{\beta\alpha},\\ g_{m\alpha}&=g_{\alpha m}=0. \end{aligned} \tag{3.8}$$

It is also useful to define a tensor $c_{\mu\nu\lambda}$ by

$$c_{\mu\nu\lambda}=g_{\mu\sigma}c^{\sigma}_{\nu\lambda}. \tag{3.9}$$

From Eq. (3.5),

$$c_{\mu\nu\lambda}=-(-1)^{\varepsilon(X_\nu)\varepsilon(X_\lambda)}c_{\mu\lambda\nu}. \tag{3.10}$$

Making use of the identities (3.6), one can show (see Appendix A) that

$$c_{\mu\nu\lambda}=-(-1)^{\varepsilon(X_\mu)\varepsilon(X_\nu)}c_{\nu\mu\lambda}. \tag{3.11}$$

From Eqs. (3.10) and (3.11) we derive another "antisymmetry" relation:

$$c_{\mu\nu\lambda}=-(-1)^{\varepsilon(X_\mu)\varepsilon(X_\nu)+\varepsilon(X_\mu)\varepsilon(X_\lambda)+\varepsilon(X_\mu)\varepsilon(X_\lambda)}c_{\lambda\nu\mu}. \tag{3.12}$$

Definition: A graded Lie algebra is called *semisimple* if

$$\det|g_{\mu\nu}|\neq 0,\quad \det|h_{mn}|\neq 0. \tag{3.13}$$

For semisimple graded Lie algebras we can define a contravariant metric tensor $g^{\mu\nu}$ through the relation

$$g_{\mu\lambda}g^{\lambda\nu}=\delta^{\nu}_{\mu}. \tag{3.14}$$

Theorem: The operator

$$K\equiv g^{\mu\nu}X_\mu X_\nu=g^{mn}Q_mQ_n+g^{\alpha\beta}V_\alpha V_\beta \tag{3.15}$$

is a Casimir operator.

Proof:

$$\begin{aligned}[X_\lambda,K]&=((-1)^{\varepsilon(X_\lambda)(\varepsilon(X_\mu)+\varepsilon(X_\nu))}-1)g^{\mu\nu}X_\mu X_\nu X_\lambda\\ &\quad+((-1)^{\varepsilon(X_\lambda)\varepsilon(X_\mu)}g^{\mu\nu}c^{\sigma}_{\lambda\nu}+c^{\mu}_{\lambda\nu}g^{\nu\sigma})X_\mu X_\sigma. \end{aligned}\tag{3.16}$$

Each of the two terms on the right-hand side of Eq. (3.16) vanishes separately,

$$\begin{aligned}[(-1)^{\varepsilon(X_\lambda)[\varepsilon(X_\mu)+\varepsilon(X_\nu)]}-1]g^{\mu\nu}&=[(-1)^{2\varepsilon(X_\lambda)\varepsilon(X_\mu)}-1]g^{\mu\nu}\\ &=0,\end{aligned}$$

and, using Eq. (3.12), we have

$$\begin{aligned}&(-1)^{\varepsilon(X_\lambda)\varepsilon(X_\mu)}g^{\mu\nu}c^{\sigma}_{\lambda\nu}+c^{\mu}_{\lambda\nu}g^{\nu\sigma}\\ &=[(-1)^{\varepsilon(X_\lambda)\varepsilon(X_\mu)}g^{\sigma\nu}c_{\nu\lambda\tau}+c_{\tau\lambda\nu}g^{\nu\sigma}]g^{\mu\tau}\\ &=[-(-1)^{\varepsilon(X_\lambda)\varepsilon(X_\mu)+\varepsilon(X_\sigma)\varepsilon(X_\lambda)+\varepsilon(X_\sigma)\varepsilon(X_\tau)+\varepsilon(X_\lambda)\varepsilon(X_\tau)}\cdot g^{\sigma\nu}\\ &\quad+g^{\nu\sigma}]g^{\mu\tau}c_{\nu\lambda\tau}\\ &=[-(-1)^{\varepsilon(X_\sigma)[\varepsilon(X_\lambda)+\varepsilon(X_\mu)+\varepsilon(X_\sigma)]}+1]\cdot g^{\nu\sigma}g^{\mu\tau}c_{\nu\lambda\tau}=0,\end{aligned}$$

since in $c_{\nu\lambda\tau}$ there are always two odd operators or none and thus

$$\left\{g(X_\sigma)[g(X_\lambda)+g(X_\mu)+g(X_\sigma)]=\begin{cases}0 & \text{if } g(X_\sigma)=0,\\ 2 & \text{if } g(X_\sigma)=1.\end{cases}\right.$$

The importance of the existence of the invariant quadratic form has been discussed in Sec. 1 and we have also seen in Sec. 2 in the example of the GSU(2) algebra that the existence of the Casimir operator (3.15) made it possible not only to classify the irreducible representations but also to prove the total reducibility of the representations.

We will prove next several theorems which will enable us to make a classification of the semisimple graded algebras and to bring them to a standard form. In order to do so, we first write Eq. (1.1) in the standard Cartan basis:

$$\begin{aligned}&[H_i,H_j]=0,\quad i,j=1,\ldots,l,\\ &[H_i,E_a]=a_iE_a,\\ &[E_a,E_{-a}]=a^iH_i,\\ &[E_a,E_b]=N_{a,b}E_{a+b}\quad(a+b\neq 0).\end{aligned}\tag{3.17}$$

l is the rank of the algebra, H_i are its center elements. The a_i are the covariant roots, the a^i their contravariant counterparts. $E_a,E_i,\cdots$ are the noncenter generators. The connection between the a^i and the a_i depends on a normalization convention which we will specify shortly.

Next we write in the same basis some equations involving the V's:

$$\begin{aligned}&[H_i,V_{(\alpha,r)}]=\alpha_iV_{(\alpha,r)},\\ &[E_a,V_{(\alpha,r)}]=\delta^{\beta}_{a+\alpha}M_{(\beta,s),(\alpha,r)}V_{(\beta,s)},\\ &\{V_{(\alpha,r)},V_{(-\alpha,s)}\}=A^{i}_{(\alpha,r),(-\alpha,s)}H_i,\\ &\{V_{(\alpha,r)},V_{(\beta,s)}\}=\delta^{\alpha}_{\alpha+\beta}A_{(\alpha,r),(\beta,s)}E_a\quad(\alpha+\beta\neq 0).\end{aligned}\tag{3.18}$$

The first two of these equations are a transcription of Eq. (1.2). The α_i are the weights. To a given weight there may correspond several V's. We express this by the degeneracy index r in $V_{(\alpha,r)}$. The second pair of Eqs. (3.18) are a consequence of Eq. (1.3) and of Eq. (1.6) with $Q_m\to H_i$.

Theorem: If the graded Lie algebra is semisimple, then to each weight α of the grading representation V, there corresponds a weight $-\alpha$.

Proof: From (3.8) and (3.18) we have

$$g_{(\alpha,r),(\beta,s)}=\delta_{\alpha,-\beta}f_{(\alpha,r),(-\alpha,s)}, \tag{3.19}$$

where

$$\begin{aligned}f_{(\alpha,r),(-\alpha,s)}&=\sum_{a,t}(A_{(-\alpha,s),(a+\alpha,t)}M_{(a+\alpha,t),(\alpha,r)}\\ &\quad-A_{(\alpha,r),(a-\alpha,t)}M_{(a-\alpha,t),(-\alpha,s)})\\ &=-f_{(-\alpha,s),(\alpha,r)}.\end{aligned}\tag{3.20}$$

The condition of semisimplicity (3.13) implies that $f_{(\alpha,r),(-\alpha,s)}\not\equiv 0$. Hence to each α corresponds a weight α and for each (α,r) there exists at least one state $(-\alpha,s)$ such that $f_{(\alpha,r),(-\alpha,s)}$ does not vanish.

The other elements of the metric tensor can also be computed, and we obtain (an i,j index refers to the H's, an a index refers to the E's);

$$g_{ij}=\sum_a a_ia_j-\sum_\alpha n(\alpha)\alpha_i\alpha_j, \tag{3.21}$$

$$g_{ia}=0,$$

$$g_{ab}=\delta_{a,-b}\Big(-2a^i a_i+\sum_c N_{ac}N_{-a,a+c}-\sum_{\alpha,r,s}M_{(a+\alpha,s),(\alpha,r)}M_{(\alpha,r),(a+\alpha,s)}\Big), \tag{3.22}$$

where $n(\alpha)$ denotes the multiplicity of the weight α. We will choose our normalization such that

$$g_{ab}=\delta_{a,-b}. \tag{3.23}$$

We will implement the content of the identities (1.5)–(1.7) in two stages. First we will impose the "anti-symmetry" constraints (3.11). Thereafter we will check the remaining constraints following from Eqs. (1.5)–(1.7).

From Eq. (3.11) we find

$$c_{iab}=\delta_{a,-b}g_{il}a^l=-c_{aib}=\delta_{a,-b}a_i. \tag{3.24}$$

This implies

$$a_i=g_{il}a^l. \tag{3.25}$$

Also from Eq. (3.11) we get

$$c_{ija}=0, \tag{3.26}$$

$$c_{abc}=\delta_{b+c,-a}N_{bc}=-c_{bac}. \tag{3.27}$$

Equation (3.27) holds if

$$N_{ab}=-N_{-(a+b),b}. \tag{3.28}$$

Again from Eq. (3.11)

$$c_{i,(\alpha,r),(\beta,s)}=\delta_{\alpha,-\beta}g_{il}A^l_{(\alpha,r),(-\alpha,s)}=-c_{(\alpha,r),i,(\beta,s)}. \tag{3.29}$$

Equation (3.29) holds if

$$A^i_{(\alpha,r),(-\alpha,s)}=f_{(\alpha,r),(-\alpha,s)}\alpha^i. \tag{3.30}$$

Once more from Eq. (3.11),

$$c_{a,(\alpha,r),(\beta,s)}=\delta^{-a}_{\alpha+\beta}A_{(\alpha,r),(-\alpha-a,s)}=-c_{(\alpha,r),a,(\beta,s)}. \tag{3.31}$$

Equation (3.31) holds if

$$A_{(\alpha,r),(\beta,s)}=-f_{(\alpha,r),(-\alpha,t)}M_{(-\alpha,t),(\beta,s)} \tag{3.32a}$$

and

$$f_{(\alpha,r),(-\alpha,t)}M_{(-\alpha,t),(\beta,s)}=f_{(\beta,s),(-\beta,u)}M_{(-\beta,u),(\alpha,r)}. \tag{3.32b}$$

We now check the Jacobi identity (1.5) according to which

$$F^{(a)}_{(\alpha,r),(\alpha,s)}-F^{(-a)}_{(\alpha,r),(\alpha,s)}=-a^i\alpha_i\delta_{rs}, \tag{3.33}$$

$$F^{(-a)}_{(\alpha,r),(\alpha-a-b,s)}-F^{(-b)}_{(\alpha,r),(\alpha-a-b,s)}=N_{ab}M_{(\alpha,r),(\alpha-a-b,s)}, \tag{3.34}$$

where

$$F^{(a)}_{(\alpha,r),(\beta,s)}=\sum_t M_{(\alpha,r),(a+\alpha,t)}M_{(a+\alpha,t),(\beta,s)}.$$

By using (3.30), (3.32)–(3.34), the identity Eq. (1.6) is automatically satisfied.

We turn now to the identities (1.7). They have the following explicit expression in the Cartan basis:

$$\begin{aligned}&\delta_{\alpha,-\beta}\delta_{ut}\alpha^i\gamma_i f_{(\alpha,r),(-\alpha,s)}+\delta_{\beta,-\gamma}\delta_{ur}\beta^i\alpha_i f_{(\beta,s),(-\beta,t)}\\&+\delta_{\gamma,-\alpha}\delta_{us}\gamma^i\beta_i f_{(\gamma,t),(-\gamma,r)}\\&-\sum_v f_{(\alpha,r),(-\alpha,v)}M_{(-\alpha,v),(\beta,s)}M_{(\alpha+\beta+\gamma,u),(\gamma,t)}\\&-\sum_v f_{(\beta,s),(-\beta,v)}M_{(-\beta,v),(\gamma,t)}M_{(\alpha+\beta+\gamma,u),(\alpha,r)}\\&-\sum_v f_{(\gamma,t),(-\gamma,v)}M_{(-\gamma,v),(\alpha,r)}M_{(\alpha+\beta+\gamma,u),(\beta,s)}\\&=0.\end{aligned} \tag{3.35}$$

It is from these last relations that we will obtain several theorems which will allow the classification of the semisimple graded algebras. In order to do so, we need the following lemma.

Lemma: If A and B are two symmetric $n\times n$ matrices acting in an n-dimensional space and if

$$A_{r_1r_2}B_{s_1s_2}=\delta_{r_1s_1}\delta_{r_2s_2}+\delta_{r_1s_2}\delta_{r_2s_1},\quad r_1,r_2,s_1,s_2=1,2,\ldots,n, \tag{3.36}$$

then the vector space is one-dimensional ($n=1$).

Proof: Assume $A_{11}\neq 0$. From (3.36) we get

$$A_{11}B_{11}=2,\quad B_{s_1s_2}=0 \text{ for } (s_1,s_2)\neq(1,1).$$

Take $(r_1,r_2)\neq(1,1)$. Then none of the relations (3.36) can be satisfied which implies $r_1=r_2=s_1=s_2=1$.

Let us next define the inverse of the matrix $f_{(\alpha,r),(-\alpha,s)}$:

$$\sum_t f^{(-1)}_{(-\alpha,r),(\alpha,t)}f_{(\alpha,t),(-\alpha,s)}=\delta_{rs}. \tag{3.37}$$

Theorem: All the nonzero weights are simple.

Proof: Take $\beta=\gamma=-\alpha$ in Eq. (3.36):

$$\alpha^i\alpha_i(f_{(-\alpha,t),(\alpha,r)}\delta_{us}+f_{(-\alpha,s),(\alpha,r)}\delta_{ut})=\sum_v f_{(-\alpha,s),(\alpha,v)}M_{(\alpha,v),(-\alpha,t)}M_{(-\alpha,u),(\alpha,r)}. \tag{3.38}$$

Multiply this relation by $f^{(-1)}_{(\alpha,r),(-\alpha,w)}$ and sum over r:

$$\alpha^i\alpha_i(\delta_{tw}\delta_{us}+\delta_{sw}\delta_{tu})=\sum_v f_{(-\alpha,s),(\alpha,v)}M_{(\alpha,v),(-\alpha,t)}\cdot\sum_r M_{(-\alpha,u),(\alpha,r)}f^{(-1)}_{(\alpha,r),(-\alpha,w)}. \tag{3.39}$$

It follows from the lemma that the space (α,r) for given α is one-dimensional and Eq. (3.39) becomes

$$2\alpha^i\alpha_i=M_{\alpha,-\alpha}M_{-\alpha,\alpha}, \tag{3.40}$$

since we now can drop the index r in (α,r).

Root-weight theorem: To each weight α corresponds a root a such that $a=2\alpha$ and

$$\{V_\alpha,V_\alpha\}\neq 0. \tag{3.41}$$

Proof: From Eq. (3.40) it follows that $M_{-\alpha,\alpha}\neq 0$. But

$$[E_{-2\alpha},V_\alpha]=M_{-\alpha,\alpha}V_{-\alpha}.$$

Hence $a=-2\alpha$ exists and so does $a=2\alpha$. Let us consider the expression

$$\{V_\alpha,V_\alpha\}=-f_{\alpha,-\alpha}M_{-\alpha,\alpha}E_{2\alpha}. \tag{3.42}$$

Since $f_{\alpha-\alpha}\neq 0$ and $M_{-\alpha,\alpha}\neq 0$, Eq. (3.41) results.

Lemma: If a is a root, ka (k integer) is a root only if $k=\pm 1$. The proof is given elsewhere.[12]

Theorem: There are no zero weights in V_α.

Proof: Let α be a weight. Then

$$\{V_{(0,s)},V_\alpha\}=-f_{\alpha,-\alpha}M_{-\alpha,(0,s)}E_\alpha.$$

But $E_{2\alpha}$ is a root so E_α cannot be a root. This implies $M_{-\alpha,(0,s)}=0$. We also have

$$[E_a, V_{(0,s)}] = M_{a,(0,s)} V_a = 0.$$

Since

$$\{V_{(0,s)}, V_{(0,r)}\} = [H_i, V_{(0,s)}] = 0,$$

we derive that [see Eq. (3.20)]

$$f_{(0,r),(0,s)} = 0.$$

The theorem then follows from the requirement of semisimplicity. Since all the weights are simple, we have the following obvious corollary: *The grading representation is irreducible.*

This is the central result of this section. Since the irreducible representations of the semisimple Lie algebras are classified, the problem of the classification of the semisimple graded Lie algebras has been reduced to a known one.

We can now put the commutation relations (3.18) into a standard form:

$$\begin{aligned} &[H_i, V_\alpha] = \alpha_i V_\alpha, \\ &[E_a, V_\alpha] = \delta_{a+\alpha} M_{\beta,\alpha} V_\beta, \end{aligned} \tag{3.43}$$

$$\begin{aligned} &\{V_\alpha, V_{-\alpha}\} = f_{\alpha,-\alpha} \alpha^i H_i, \\ &\{V_\alpha, V_\beta\} = -\delta_{\alpha+\beta} f_{\alpha,-\alpha} M_{-\alpha,\beta} \end{aligned} \tag{3.44}$$

where

$$f_{\alpha,-\alpha} M_{-\alpha,\beta} = f_{\beta,-\beta} M_{-\beta,\alpha}, \tag{3.45}$$

$$f_{\alpha,-\alpha} = -f_{-\alpha,\alpha}. \tag{3.46}$$

Going back to the notations of Sec. 1, we have

$$(F_m)_{\alpha\beta} = F^\alpha_{m\beta}, \quad m = 1, \ldots, D, \tag{3.47}$$

where

$$\begin{aligned} &(F_i)_{\alpha\beta} = \delta_{\alpha\beta} \alpha_i, \\ &(F_a)_{\alpha\beta} = \delta^a_{\beta-\alpha} M_{\alpha\beta}. \end{aligned} \tag{3.48}$$

If we define the matrix C though the relation

$$C_{\alpha\beta} = -\delta_{\alpha,-\beta} f_{\alpha,-\alpha}, \tag{3.49}$$

we have

$$C = -C^t, \quad (CF_m) = (CF_m)^t. \tag{3.50}$$

If we choose our coordinates such that $g_{ij} = \delta_{ij}$, we also have $\alpha_i = \alpha^i$ and from (3.43):

$$A^k = CF_k \tag{3.51}$$

We therefore have the *C-theorem*: For each SSGLA there exists a matrix C which satisfies Eq. (3.50), (3.51).

A representation which is real and satisfies the condition (3.50) is called a real negative representation[10] (a representation is called real positive if $C = C^t$).

We can now summarize our results as follows: *The grading representation of a SSGLA is irreducible, real negative and to each weight α corresponds a root $a = 2\alpha$.*

4. IMPLICATIONS OF THE ROOT-WEIGHT THEOREM

In the previous section we found two necessary conditions for a Lie algebra to be gradable. First, it must contain at least one irreducible representation which is real and negatvie (RNR). Secondly, if such a representation exists then to each of its weights must correspond a root of the algebra which equals twice that weight.

In this section we shall show that these two criteria suffice to rule out all the algebras SU(N), $N > 2$, all O(N), $N > 5$, and all the exceptional algebras as candidates for gradability. We shall show further that for the symplectic algebras Sp($2N$) there is one and only one candidate for the grading representation V_α, namely the fundamental $2N$-dimensional representation. For O(3), O(4), O(5) there is also only one candidate, namely the spinor representation. Since the algebras of SU(2), O(3), and Sp(2) are isomorphic and have a common two-dimensional representation, since O(4) is isomorphic to Sp(2)×Sp(2) with a shared four-dimensional representation, and since the same is true for O(5) and Sp(4), we conclude that only the algebras Sp($2N$) are candidates for grading and that they are so in a unique way. In the next section we shall show that for these cases the grading can actually be implemented.

In order to proceed, we make use of the results of Ref. 10. There it was shown for each SSLA which representations are real and negative. The same papers also contain expressions for the root vectors and the dominant weights in terms of an orthogonal basis. We shall see next that the results just mentioned are obtainable if we confine our attention to the highest weight only and seek for a corresponding root which satisfies our root-weight theorem.

The highest weight M of an irreducible representation characterizes the representation and can be written in terms of l fundamental dominant weights Π_j (l is the rank of the algebra)

$$M = \sum \lambda_i \Pi_i, \tag{4.1a}$$

with nonnegative λ_i. The Π_i are tabulated in Ref. 10.

Thus we hope to find a set λ_i and a root a such that

$$a = 2M. \tag{4.1b}$$

The algebras A_{n-1} [$SU(N)$]: These contain RNR only if $n = 4p + 2$ and the λ_i satisfy the conditions:

$$\lambda_i = \lambda_{n-i} \quad (i = 1, \ldots, n-1), \tag{4.2}$$

and $\lambda_{n/2}$ is odd.

The roots can be expressed in terms of fundamental weights as follows:

$$\pm a(j,k) = \Pi_{n-k+1} - \Pi_{n-k+2} - \Pi_{n-j+1} + \Pi_{n-j+2} + \delta_{1,j}\Pi_1, \quad (1 \le j < k \le n). \tag{4.3}$$

The conditions (4.1) can be satisfied only if $n = 2$ when

$$\pm a(1,2) = 2\Pi_1,$$

when indeed $\lambda_1 = 1$ is odd.

The algebras B_n [$O(2n+1)$] contain RNR only if $n(n+1)/2$ is an odd number and if $\lambda_1 \neq 0$. We define the vectors P_i through the relations

$$P_1 = \Pi_2, \quad \cdots, \quad P_{n-1} = \Pi_n, \quad P_n = 2\Pi_1, \tag{4.4}$$

and express the roots in terms of the P_i:

$$\pm a(\pm j) = P_j - P_{j-1}, \tag{4.5a}$$

$$\pm a(\pm j, \pm k) = P_j - P_{j-1} + P_k - P_{k-1}, \tag{4.5b}$$

$$\pm a(\pm j, \mp k) = P_j - P_{j-1} - P_k + P_{k-1}. \tag{4.5c}$$

The conditions (4.1) can be satisfied by the roots (4.5a) only if $n=1$, in which case

$$\pm a(\pm 1) = 2\Pi_1. \tag{4.6}$$

The roots (4.5b) check (4.1) only if $n=2$ and then

$$\pm a(\pm 1, \pm 2) = 2\Pi_1. \tag{4.7}$$

For both $n=1$ and 2, $\lambda_1 = 1$ and $n(n+1)/2$ is odd.

The algebras C_n $[Sp(2n)]$ contain RNR only if $\lambda_1 + \lambda_3 + \lambda_5 + \cdots$ is odd. The roots are

$$\pm a(\pm j) = 2(\Pi_j - \Pi_{j-1}), \tag{4.8a}$$

$$\pm a(\pm j, \pm k) = \Pi_j - \Pi_{j-1} + \Pi_k - \Pi_{k-1}, \tag{4.8b}$$

$$\pm a(\pm j, \mp k) = \Pi_j - \Pi_{j-1} - \Pi_k + \Pi_{k-1}. \tag{4.8c}$$

In this case

$$\pm a(\pm 1) = 2\Pi_1.$$

Thus $\lambda_1 = 1$, $\lambda_i = 0$ $(i = 2, \ldots, n)$, and $\lambda_1 + \lambda_3 + \lambda_5 + \cdots = \lambda_1$ is odd. The dimension of the representation having the highest weight Π_1 is[10] $2n$.

The algebras D_n $[O(2n)]$: We define the vectors P_j in terms of the fundamental weights Π_j:

$$P_j = \Pi_{j+2} \quad (j = 1, \ldots, n-2), \quad P_{n-1} = \Pi_1 + \Pi_2, \quad P_n = 2\Pi_1. \tag{4.9}$$

The roots are

$$\pm a(\pm j, \pm k) = P_j - P_{j-1} + P_k - P_{k-1}, \tag{4.10a}$$

$$\pm a(\pm j, \mp k) = P_j - P_{j-1} - P_k + P_{k-1}. \tag{4.10b}$$

The conditions (4.1) are satisfied only for $n=2$, $j=1$, $k=2$ in which case

$$\pm a(\pm 1, \pm 2) = 2\Pi_1.$$

This implies $\lambda_1 = 1$, $\lambda_i = 0$ $(i = 2, \ldots, n)$ and thus $\lambda_1 + \lambda_2$ is odd.

The algebras G_2, F_4, E_6, and E_8: These have no RNR.

The algebra E_7: has 126 roots which can be expressed in terms of the fundamental weights. We verified that none of these roots satisfy the condition (4.1), and thus the algebra E_7 cannot be graded.

We conclude this section with a further elaboration of Eq. (4.1b). In that equation M is the *highest* weight vector of the grading representation, as already noted. Furthermore, it is a consequence of our analysis of gradable algebras that the root vector a which appears in Eq. (4.1b) is specifically the *highest* root vector. Let us then rewrite the vector relation Eq. (4.1b) as follows.

$$M_A = 2M_V, \tag{4.11}$$

where the subscripts A and V refer to the adjoint and the grading representation, respectively. We further observe that the (length)^2 of M_V is given by

$$M_V^2 = 1, \tag{4.12}$$

if we use the normalization conventions of Racah.[9]

[*Example*: With the norms used here, we have for SU(2): $Q_m \equiv \tau_m$ for the spinor representation.]

Let $c_2(M)$ denote the eigenvalue of the quadratic Casimir operator for a representation with the highest weight M. Then[9]

$$c_2(M) = (M+R)^2 - R^2, \tag{4.13}$$

where the vector R is given by

$$R = \tfrac{1}{2} \sum_{a^+} a.$$

Here a is a root and $\sum_{a^+}$ denotes summation over positive roots only. We thus obtain

$$c_2(A) - c_2(V) = (2M_V + R)^2 - R^2 - 2(M_V + R)^2 + 2R^2 = 2,$$

which is the result stated in Eq. (1.15).

5. THE GRADING OF O(N), N = 3,4,5, AND OF Sp(2N)

These are the groups which survived after the necessary conditions for grading were implemented in Secs. 3 and 4. We shall now show that the grading of O(N), $N=3,4,5$, by spinor and of Sp($2N$) by the fundamental representation is actually possible. As noted in Sec. 1, the orthogonal cases are linked to Sp(2) and Sp(4).

We show that for the instances at hand the identities Eqs. (1.8) and (1.9) can be satisfied. We recopy these relations:

$$A^m F_n + F_n^t A^m = f_{np}^m A^p, \tag{5.1}$$

$$A_{\alpha\beta}^p F_{p\gamma}^\delta + A_{\gamma\alpha}^p F_{p\beta}^\delta + A_{\beta\gamma}^p F_{p\alpha}^\delta = 0. \tag{5.2}$$

According to Eq. (3.51), A^k is related to F_k by

$$A^k = CF_k = (CF_k)^t, \quad C^t = -C, \tag{5.3}$$

and C has an inverse. Furthermore, Eq. (5.3) corresponds to a metric where $f_{ik}^j = f_{ikj} = -f_{ijk}$. Then Eq. (5.3) inserted in Eq. (5.1) yields

$$[F_i, F_j] = f_{ijk} F_k, \tag{5.4}$$

which is identically satisfied, in accordance with the general results of Sec. 3.

We turn to Eq. (5.2). Let X_δ^6 be the most general $d \times d$ matrix (d = dimension of the grading representation). Then Eq. (5.2) is valid if for any X

$$(CF_k X F_k) + (CF_k X F_k)^t = -(CF_k)\mathrm{Tr}(XF_k). \tag{5.5}$$

Eq. (5.5) has to be verified for a complete set of d^2 linearly independent matrices X. A subset thereof may be chosen to be the generators F_i themselves. For this subset Eq. (5.5) becomes [use Eq. (5.3)]

$$2F_k F_i F_k = -F_k \,\mathrm{Tr}(F_i F_k). \tag{5.6}$$

Use Eq. (5.4) and recall that f_{ijk} is the (jk)-matrix element of F_i inasfar as F_i acts on the adjoint representation. Then Eq. (5.6) becomes

$$[c_2(A) - 2c_2(V)]F_i = F_k \,\mathrm{Tr} F_i F_k. \tag{5.7}$$

$c_2(A)$ and $c_2(V)$ are the values of the quadratic Casimir in the adjoint representation and the V representation, respectively, of the semisimple (nongraded) Lie group.

The groups $Sp(2N)$: Sp($2N$) is defined as the group which leaves invariant the form $x_i C_{ij} y_j$, $i,j = 1, \ldots, 2N$, where $C^t = -C$. The x_i and y_i are fundamental $2N$-dimen-

sional representations. C may be chosen to be in the canonical form

$$C = i\tau_2 \otimes 1, \tag{5.8}$$

where τ_2 is the usual 2×2 Pauli matrix and "1" is a $N\times N$ unit matrix. We recall that the generators F_k inasfar as they act on the fundamental representation satisfy

$$CF_k + F_k^t C = 0. \tag{5.9}$$

A complete set of $N(2N+1)$ such matrices F_k can be constructed as follows. Let λ_α denote the set of $N\times N$ matrices corresponding to the SU(N) generators acting on a N-dimensional representation of that group. Divide this set of N^2-1 matrices into two subsets $\lambda_\alpha^{(s)}$ and $\lambda_\alpha^{(a)}$, where the superscripts (s) or (a) denote that the λ_α in question is symmetric or antisymmetric respectively. Then the collection of F_k matrices in Eq. (5.9) is easily seen to be given by

$$F_k\colon\ \tau_i \otimes 1,\quad 1 \otimes \lambda_\alpha^{(a)},\quad \tau_i \otimes \lambda_\alpha^{(s)}. \tag{5.10}$$

For the present case the complete set of matrices X which enter into Eq. (5.5) is given by

$$X\colon\ 1 \otimes 1,\quad \tau_i \otimes 1,\quad 1 \otimes \lambda_\alpha,\quad \tau_i \otimes \lambda_\alpha. \tag{5.11}$$

For definiteness, the λ_α are normed in such a way that for each pair X_A, X_B of X's

$$\mathrm{Tr} X_A X_B = 2N\delta_{AB}. \tag{5.12}$$

The X-set contains the F_k as a subset. Let X'_A denote any X matrix not in this subset. One easily verifies that

$$CX'_A - X'^t_A C = 0. \tag{5.13}$$

We first verify Eq. (5.7). With our norm convention one finds[13] $c_2(V) = N(2N+1)$, $c_2(A) = 4N(N+1)$. From this and from Eq. (5.12), Eq. (5.7) is seen to be satisfied.

For any X'_A, the left-hand side of Eq. (5.5) vanishes because of Eq. (5.13), the antisymmetry of C, and the symmetry of CF_k. The right-hand side of Eq. (5.5) also vanishes because of Eq. (5.12). Thus all identities are satisfied.

The groups $O(3)$, $O(4)$, $O(5)$: From Eq. (5.10) we see that $SU(2) \approx Sp(2)$. For Sp(4), Eq. (5.10) gives $\tau_i \otimes 1$, $\tau_i \otimes \eta_1$, $\tau_i \otimes \eta_3$ and $1 \otimes \eta_2$, for the set of generators, where the η_i are a second set of Pauli matrices. These are the ten generators acting on an O(5) spinor and satisfying the O(5) algebra. For O(4) one can take again $C = i\tau_2 \otimes 1$. The set of generators is $\tau_i \otimes 1$, $\tau_i \otimes \eta_1$ and for the X' one has the remaining ten Dirac matrices. The verification of the identities goes as before.

Finally we note that the spinor representations of $O(N)$ provide an illuminating example of the fact that the existence of a C matrix such that $C^t = -C$, $(CF_k)^t = (CF_k)$ is not sufficient to effect grading. Namely, it has been shown[14] that for these spinor representations there does exist a C matrix with these properties for $N = 3, 4, 5, 6$ (mod 8). For all such cases one can proceed to construct a set of X matrices along similar lines as spelled out above. It can then be shown that Eq. (5.7) leads to the constraint

$$N^2 - 9N + 16 = -2^{[N/2]}, \tag{5.14}$$

where $[\chi]$ is the entire part of χ. This equation has as its only solutions $N = 3, 4, 5$—as it should be.

6. NONSEMISIMPLE GRADING: SU(N), $N>2$

These SU(N) groups provide examples of simple Lie algebras whose grading is nonsemisimple. To start with, consider any Lie algebra and ask whether it can be graded by its adjoint representation. In that case, the F_m in Eq. (1.2) refer to the adjoint representation and[7]

$$(F_n)^p_m = f^p_{nm}. \tag{6.1}$$

Then it follows from Eq. (3.8) that $g_{mn} \equiv 0$. Since $g_{m\alpha}$ and $g_{\alpha n}$ always vanish [see Eq. (3.8)], we have from Eq. (3.13): Any possible grading of a Lie algebra by its adjoint representation is necessarily nonsemisimple.

Furthermore, for the adjoint representation we always have

$$F_n = -F_n^t, \tag{6.2}$$

so that Eq. (1.8) takes the form

$$[F_n, A^m] = -f^m_{np}A^p = -f_{npm}A^p = f_{nmp}A^p. \tag{6.3}$$

(Here we use the conventional metric.) For any Lie algebra, Eq. (6.3) by itself is satisfied by $A^m = F_m$, but this solution is unacceptable because Eq. (1.4) conflicts with Eq. (6.2). It is at this point that we turn specifically to the groups SU(N), $N>2$.

As is well known, for these groups there is a second solution to Eq. (6.3), namely

$$A^m = D^m,\quad (D^m)_{kl} = d_{mkl}, \tag{6.4}$$

where d_{mkl} is totally symmetric, so that Eq. (1.4) is now satisfied. The d symbols are defined as follows. Let $F^m_{(N)}$ be the set of generators of SU(N) inasfar as they act on the fundamental representation (N). Then, for any N,

$$\{F^i_{(N)}, F^j_{(N)}\} = d_{ijk}F^k_{(N)} + (1/N)\delta_{ij}. \tag{6.5}$$

(Here, for definiteness, a normalization has been chosen such that $\mathrm{Tr}(F^i_{(N)})^2 = \frac{1}{2}$, for all N, fixed i). Note that $d_{ijk} \equiv 0$ for $N = 2$.

Thus our system (1.1)–(1.3) now reads

$$\begin{aligned} [Q_m, Q_n] &= f^p_{mn}Q_p, \\ [Q_m, V_n] &= f^p_{mn}V_p, \\ \{V_m, V_n\} &= d_{mnp}Q_p. \end{aligned} \tag{6.6}$$

One easily verifies that Eq. (1.7) is also satisfied, since

$$d_{mnp}f_{pqr} + d_{qmp}f_{pnr} + d_{nqp}f_{pmr} = 0. \tag{6.7}$$

Returning once more to Eq. (3.8), we see that in the present case $g_{\alpha\beta}$ also vanishes, as follows from the relation $F_m D_m = D_m F_m = 0$. Therefore, this grading is not only nonsemisimple, but we have the even more singular situation that the metric tensor $g_{\mu\nu}$ vanishes identically.

Of course, there is here no such theorem as the one related to Eq. (3.15). Yet there does exist a quadratic operator K_2 which satisfies Eq. (1.11), namely

$$K_2 = Q_m^2. \tag{6.8}$$

Indeed, it is evident from Eq. (6.6) that $[Q_m^2, Q_n] = [Q_m^2, V_n] = 0$. Note also that $V_m^2 \equiv 0$.

As an example of an explicit representation of the algebra we merely mention the $(2N^2 - 2)$-dimensional representation

$$\begin{pmatrix} A \\ A' \end{pmatrix}, \tag{6.9}$$

where both A and A' are $(N^2 - 1)$-dimensional adjoint representations. The realization of the operators Q_m, V_m is as follows for this case.

$$Q_m = \begin{pmatrix} F_m & 0 \\ 0 & F_m \end{pmatrix}, \quad V_m = \begin{pmatrix} 0 & F_m \\ D_m & 0 \end{pmatrix}. \tag{6.10}$$

This is readily verified with the help of Eq. (6.7).

7. FURTHER QUESTIONS

It will be clear that a considerable amount of work still needs to be done in order to map fully the new structures which have been uncovered in the foregoing. We conclude this paper by listing some questions which seem to us to be most interesting.

1) As already stated in Sec. 1, one would like to know whether or not there exist graded Lie algebras with $\det|g_{\mu\nu}| \neq 0$, $\det|h_{mn}| = 0$.

2) For all GSp($2N$) one ought to know whether the representations are fully reducible (we guess that they are).

3) Related to the preceding question is the problem of the existence of Casimir operators K defined by Eq. (1.11) which are of degree higher than two in the Q_m and the V_α.

4) A general proof of the Wigner–Eckart theorem is still outstanding, see Sec. 2.

5) Just as we did for GSU(2), one should not only examine the algebra of GSP($2N$), $N > 1$, but also the corresponding group structure.

APPENDIX A

We derive here Eq. (3.11). In order to do so, we write the identities (1.5)–(1.7) along with the Jacobi identities corresponding to the Lie algebra (1.1):

$$f^p_{nr} f^q_{mp} + f^p_{rm} f^q_{np} + f^p_{mn} f^q_{rp} = 0, \tag{A1}$$

$$f^p_{mn} F^\delta_{p\alpha} - F^\gamma_{n\alpha} F^\delta_{m\gamma} + F^\gamma_{m\alpha} F^\delta_{n\gamma} = 0, \tag{A2}$$

$$A^p_{\beta\gamma} f^n_{mp} - F^\delta_{m\gamma} A^n_{\beta\delta} - F^\delta_{m\beta} A^n_{\gamma\delta} = 0, \tag{A3}$$

$$A^p_{\beta\gamma} F^\delta_{p\alpha} + A^p_{\gamma\alpha} F^\delta_{p\beta} + A^p_{\alpha\beta} F^\delta_{p\gamma} = 0. \tag{A4}$$

From Eq. (3.9) we get

$$c_{nrs} = (f^q_{pn} f^p_{qm} - F^\beta_{n\alpha} F^\alpha_{m\beta}) f^m_{rs}, \tag{A5}$$

$$c_{n\alpha\beta} = (f^q_{pn} f^p_{qm} - F^\gamma_{n\delta} F^\delta_{m\gamma}) A^m_{\alpha\beta}, \tag{A6}$$

$$c_{\alpha r\beta} = (F^\gamma_{p\alpha} A^p_{\gamma\delta} - F^\gamma_{p\delta} A^p_{\gamma\alpha}) F^\delta_{n\beta}. \tag{A7}$$

To prove that

$$c_{nrs} = -c_{rns}, \tag{A8}$$

we note that from Eq. (A1) we have

$$f^q_{mr} f^m_{qp} f^p_{ns} = -f^q_{mn} f^m_{qp} f^p_{rs}. \tag{A9}$$

Using Eq. (A2), we obtain

$$F^\alpha_{m\beta} f^n_{rs} = F^\gamma_{s\beta} F^\alpha_{r\gamma} - F^\gamma_{r\beta} F^\alpha_{s\gamma} \tag{A10}$$

and

$$-F^\beta_{n\alpha} F^\alpha_{m\beta} f^m_{rs} = F^\gamma_{s\beta} (F^\alpha_{m\gamma} F^\beta_{r\alpha} - F^\beta_{n\alpha} F^\alpha_{r\gamma}). \tag{A11}$$

From (A5), (A9), and (A11) we get (A8).

We now prove

$$c_{n\alpha\beta} = -c_{\alpha n\beta}. \tag{A12}$$

From Eq. (A3) we have

$$f^q_{pn} f^p_{qm} A^m_{\alpha\beta} = F^\delta_{r\alpha} A^p_{\beta\delta} f^r_{pn} + F^\delta_{p\beta} A^r_{\alpha\delta} f^p_{rn}, \tag{A13}$$

and from Eq. (A4) we obtain

$$F^\gamma_{n\delta} F^\delta_{r\gamma} A^r_{\alpha\beta} = -F^\gamma_{n\delta} A^p_{\beta\gamma} F^\delta_{p\alpha} - F^\gamma_{n\delta} A^p_{\gamma\alpha} F^\delta_{p\beta}. \tag{A14}$$

By using (A13) and (A14), Eq. (A5) becomes

$$c_{n\alpha\beta} = F^\gamma_{n\delta} F^\delta_{p\alpha} A^p_{\beta\gamma} + F^\gamma_{n\delta} A^p_{\gamma\alpha} F^\delta_{p\beta} + f^p_{qn} F^\delta_{p\alpha} A^q_{\beta\delta} + f^p_{qn} A^q_{\alpha\delta} F^\delta_{p\beta}. \tag{A15}$$

From Eq. (A3) we get

$$F^\gamma_{p\alpha} A^p_{\gamma\delta} F^\delta_{n\beta} = F^\gamma_{p\alpha} A^q_{\beta\gamma} f^p_{nq} - F^\gamma_{p\alpha} F^\delta_{n\gamma} A^p_{\beta\delta}, \tag{A16}$$

and from Eq. (A2) we have

$$A^p_{\gamma\alpha} F^\gamma_{p\delta} F^\delta_{n\beta} = F^\delta_{p\beta} A^p_{\gamma\alpha} F^\gamma_{n\delta} + A^p_{\gamma\alpha} f^q_{pn} F^\gamma_{q\beta}. \tag{A17}$$

By using (A16) and (A17), Eq. (A7) becomes

$$c_{\alpha n\beta} = -F^\delta_{n\gamma} F^\gamma_{p\alpha} A^p_{\beta\delta} - F^\gamma_{n\delta} A^p_{\gamma\alpha} F^\delta_{p\beta} + f^p_{nq} F^\gamma_{p\alpha} A^q_{\beta\gamma} - f^p_{qn} A^q_{\gamma\alpha} F^\gamma_{p\beta}. \tag{A18}$$

Comparing (A15) with (A18) we obtain Eq. (A12).

Equations (A8) and (A12) can be written in a compact form which is (3.11).

APPENDIX B

Examples of graded Lie algebras for which $\det|g_{\mu\nu}| = \det|h_{mn}| = 0$ are legion. The simplest example is obtained by a grading of U(1)×U(1). There are two generators Y and Z, $[Y, Z] = 0$. The grading representation is given by the pair V, W which satisfies

$$[Y, V] = V, \quad [Y, W] = -W, \quad [Z, V] = [Z, W] = 0,$$
$$V^2 = W^2 = 0, \quad \{V, W\} = Z.$$

It is obvious that the determinants of $g_{\mu\nu}$ and of h_{mn} vanish and that the Eqs. (1.6) and (1.7) are satisfied.

*Work supported in part in part by the U.S. Energy Research and Development Administration under Contract Number At(11-1)-2232B.

[1]See M. Gerstenhaber, Ann. Math. **78**, 267 (1963); **79**, 59 (1964), and A. Nijenhuis and R.W.Richardson, Bull. Am. Math. Soc. 70, 406 (1964) for mathematical applications.
[2]F.A. Berezin and G.I. Kac, Mat. Sbornik **82**, 124 (1970) [Math. USSR Sb. **11**, 311 (1970)].
[3]P. Ramond, Phys. Rev. D **3**, 2415 (1971).
[4]A. Neveu and J.H. Schwarz, Nucl. Phys. B **31**, 86 (1971).
[5]J. Wess and B. Zumino, Nucl. Phys. B **70**, 39 (1974). See also D.V. Volkov and V.P. Akulov, Phys. Lett. **46B**, 109 (1973).

[6]For a recent review and detailed references, see B. Zumino, Proc. XVII Intern. Conf. on High Energy Physics, I-245 (1974), Publ. The Research Council, Rutherford Laboratory. See further R. Haag, J.T. Lopuszanski and M. Solmius, Nucl. Phys. B 88, 257 (1975).

[7]$F^{\gamma}_{m\alpha}$ is considered as a matrix F_m in which γ labels the rows, α the columns.

[8]SSGLA = semisimple graded Lie algebra; SSLA = semisimple Lie algebra.

[9]G. Racah, "Group Theory and Nuclear Spectroscopy", Princeton 1951, unpublished mimeographed notes.

[10]M. L. Mehta, J. Math. Phys. 7, 1824 (1966); M. L. Mehta and P.K. Srivastava, J. Math. Phys. 7, 1833 (1966).

[11]L. Corwin, Y. Ne'eman, and S. Sternberg, Rev. Mod. Phys. 47, 573 (1975).

[12]B.G. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1974), p. 64.

[13]For explicit expressions of the c_2's see A.M. Perelomov and V.S. Popov, Yad. Fiz. 3, 1127 (1966); [Sov. J. Nucl. Phys. 3, 819 (1966)].

[14]A. Pais, J. Math. Phys. 3, 1135 (1962).

Transport equations for the Stokes parameters from Maxwell's equations in a random medium*

G. C. Papanicolaou and R. Burridge

Courant Institute of Mathematical Sciences, New York University, New York, New York 10012
(Received 14 April 1975)

Beginning with Maxwell's equations in a random medium and following a perturbation procedure, we obtain transport equations for the Stokes parameters. We compare our equation with Chandrasekhar's transport equation and find that they agree when the random medium is specialized appropriately. We also examine the role of degeneracy in the perturbation analysis.

1. INTRODUCTION

Our objective is to analyze the connection between radiative transport theory and the asymptotics of stochastic wave equations. One expects, from physical considerations, in the asymptotic limit of weakly inhomogeneous media with propagation over several correlation lengths and wavelengths small compared with the correlation length, that the appropriate way to describe the transport of field energy is through transport theory. Beginning with Maxwell's equations in a random medium and by following a perturbation procedure for stochastic equations, we obtain transport equations for wave amplitudes (the Stokes parameters or the coherence matrix) in the above mentioned asymptotic limit. By specializing the random medium appropriately we recover Chandrasekhar's transport equations[1] for the Stokes parameters.

In Sec. 2 we discuss the role of degeneracy in the perturbation analysis of linear stochastic equations, since Maxwell's equations are degenerate, and review the basic elements of the perturbation analysis (cf. Ref. 2 and the references cited therein). We give two examples, other than Maxwell's equations, that illustrate the role of degeneracy in determining the form of the ensuing transport equation.

In Sec. 3 we begin with the analysis of Maxwell's equations by transforming them to a form appropriate for the perturbation analysis. In Sec. 4 we apply formally the perturbation procedure of Sec. 2 and arrive at transport equations for the coherence matrix of the wave amplitudes which is simply related to the Stokes parameters. In Sec. 5 we specialize the transport equations appropriately and recover Chandrasekhar's equation. It is interesting to note that for more general random media additional terms appear in the transport equations which are not present in Chandrasekhar's equations.

The derivation of transport equations from stochastic wave equations has received considerable attention (cf. Refs. 3–12 and the references cited therein). In particular, Bourret[8,9] considers problems associated with polarization as we do here in Secs. 3–5. Aside from the formal character of our results, we also do not obtain transport equations in their most general form because we deal with statistically homogeneous random media. The extension of the present analysis to locally statistically homogeneous random media, which lead to general transport equations, requires additional considerations.

2. DEGENERATE AND NONDEGENERATE MATRIX PROBLEMS

We shall analyze a simple class of problems in order to illustrate the methods we shall employ and the form of the results that arise. We shall pay particular attention to the role played by degeneracy in the perturbation analysis.

Let $v(t)$ be a complex-valued n-dimensional vector function of time satisfying the following system of linear stochastic equations:

$$\frac{dv_p(t)}{dt} = ik_p v_p(t) + \epsilon \sum_{q=1}^{n} \mu_{pq}(t) v_q(t) + \epsilon^2 \sum_{q=1}^{n} \nu_{pq}(t) v_q(t), \quad t > 0,$$

$$v_p(0) = u_{p,0}, \quad p = 1, 2, \ldots, n. \tag{2.1}$$

The coefficients $\mu_{pq}(t)$ and $\nu_{pq}(t)$, $p, q = 1, 2, \ldots, n$, are complex valued stationary processes. We assume that their means are given by

$$E\{\mu_{pq}(t)\} = 0, \quad E\{\nu_{pq}(t)\} = \bar{\nu}_{pq}, \tag{2.2}$$

where $\bar{\nu}_{pq}$, $p, q = 1, 2, \ldots, n$, are constants, and their covariances are defined as follows:

$$\begin{aligned} &E\{\mu_{pq}(t+s)\mu_{p'q'}(s)\} = R^{\cdot,\cdot}_{pq,p'q'}(t), \\ &E\{\mu^*_{pq}(t+s)\mu_{p'q'}(s)\} = R^{*,\cdot}_{pq,p'q'}(t), \\ &E\{\mu_{pq}(t+s)\mu^*_{p'q'}(s)\} = R^{\cdot,*}_{pq,p'q'}(t), \\ &E\{\mu^*_{pq}(t+s)\mu^*_{p'q'}(s)\} = R^{*,*}_{pq,p'q'}(t), \\ &\quad p, q, p', q' = 1, 2, \ldots, n. \end{aligned} \tag{2.2'}$$

Here $E\{\cdot\}$ denotes mathematical expectation and * denotes complex conjugate. Note that the correlation functions satisfy

$$R^{*,\cdot}_{pq,p'q'}(-t) = R^{\cdot,*}_{p'q',pq}(t), \text{ etc.} \tag{2.3}$$

The correlation functions of $\nu_{pq}(t)$ will not be used in the asymptotic analysis and so they are not introduced.

The numbers k_p, $p = 1, 2, \ldots, n$, are real. When they are distinct along with their sums and differences we shall call (2.1) a nondegenerate problem. Otherwise we call it degenerate. The dimensionless parameter $\epsilon > 0$ is a measure of the size of the fluctuations. We assume that the noise intensities (cross power spectra at zero frequency)

$$\int_{-\infty}^{\infty} R^{\cdot,\cdot}_{pq,p'q'}(s)\,ds, \text{ etc.}, \quad p, q, p'q' = 1, 2, \ldots, n, \tag{2.4}$$

are finite and the integrals converge absolutely. These integrals have the dimension of time and can be thought of, alternatively, as giving a measure of the correlation time (or length) of the fluctuations. The coefficients $\nu_{pq}(t)$ are assumed to have similar properties but, because they enter as $O(\epsilon^2)$ terms in (2.1), they play a less important role in the asymptotic analysis.

Equations (2.1) arise frequently in connection with eigenfunction expansions for random wave propagation problems[10] where t plays the role of the spatial variable in the direction of propagation of the free wave. The initial value problem (2.1) corresponds to the forward scattering or parabolic approximation. The same equations (2.1) arise in quantum mechanics when the Hamiltonian has a random time-dependent perturbation [cf. Eq. (2.20) and Eq. (2.37) also].

We are interested in the behavior of the statistics of the solution $v_p(t)$, $p=1,\ldots,n$ of (2.1) when ϵ is small. Stochastic effects become significant at times of order $1/\epsilon^2$, i.e., after several correlation times. We shall explain briefly why this is so.

Let

$$\tau=\epsilon^2 t \tag{2.5}$$

and set

$$v^\epsilon(\tau)=v(\tau/\epsilon^2). \tag{2.6}$$

From (2.1) we obtain the following equation for $v^\epsilon(\tau)$.

$$\frac{dv_p^\epsilon(\tau)}{d\tau}=\frac{ik_p}{\epsilon^2}\,v_p^\epsilon(\tau)+\sum_{q=1}^{n}\frac{1}{\epsilon}\,\mu_{pq}\left(\frac{\tau}{\epsilon^2}\right)v_q^\epsilon(\tau)+\sum_{q=1}^{n}\nu_{pq}\left(\frac{\tau}{\epsilon^2}\right)v_q^\epsilon(\tau),$$
$$v_p^\epsilon(0)=u_{p,0},\quad p=1,2,\ldots,n. \tag{2.7}$$

Assume temporarily that the processes $\mu_{pq}(\tau)$ are real and let

$$\mu_{pq}^\epsilon(\tau)=\frac{1}{\epsilon}\,\mu_{pq}\left(\frac{\tau}{\epsilon^2}\right). \tag{2.8}$$

We have

$$\begin{aligned}R_{pq,p'q'}^\epsilon(\tau)&=E\{\mu_{pq}^\epsilon(\tau+\sigma)\mu_{p'q'}^\epsilon(\sigma)\}\\&=\frac{1}{\epsilon^2}E\left\{\mu_{pq}\left(\frac{\tau+\sigma}{\epsilon^2}\right)\mu_{p'q'}\left(\frac{\sigma}{\epsilon^2}\right)\right\}\\&=\frac{1}{\epsilon^2}R_{pq,p'q'}\left(\frac{\tau}{\epsilon^2}\right).\end{aligned} \tag{2.9}$$

Thus,

$$\lim_{\epsilon\downarrow 0}R_{pq,p'q'}^\epsilon(\tau)=\delta(\tau)\int_{-\infty}^{\infty}R_{pq,p'q'}(s)\,ds, \tag{2.10}$$

which means, roughly, that $\mu_{pq}^\epsilon(\tau)$ tends to white noise with noise intensity the coefficient of the delta function in (2.10). The scaling (2.5) is the only one that will do this or, in other words, we have sought the white noise limit for (2.1) because in this limit stochastic effects become most prominent.

We now return to the systematic analysis of (2.7). When $k_p=0$, $p=1,2,\ldots,n$, we have a fully degenerate problem and the limit as $\epsilon\downarrow 0$ in (2.7) is referred to as the white noise or the diffusion limit. In general, the term $ik_pv_p^\epsilon/\epsilon^2$ in (2.7) must be removed before taking the limit $\epsilon\downarrow 0$. Therefore we pass to the interaction representation and define

$$u_p^\epsilon(\tau)=\exp(-ik_p\tau/\epsilon^2)v_p^\epsilon(\tau),\quad p=1,\ldots,n, \tag{2.11}$$

which satisfies the equation

$$\begin{aligned}\frac{du_p^\epsilon(\tau)}{d\tau}&=\sum_{q=1}^{n}\exp(-ik_p\tau/\epsilon^2)\frac{1}{\epsilon}\,\mu_{pq}\left(\frac{\tau}{\epsilon^2}\right)\exp(ik_q\tau/\epsilon^2)u_q^\epsilon(\tau)\\&\quad+\sum_{q=1}^{n}\exp(-ik_p\tau/\epsilon^2)\nu_{pq}\left(\frac{\tau}{\epsilon^2}\right)\exp(ik_q\tau/\epsilon^2)u_q^\epsilon(\tau),\\u_p^\epsilon(0)&=u_{p,0},\quad p=1,2,\ldots,n.\end{aligned} \tag{2.12}$$

We may call the limit $\epsilon\downarrow 0$ in (2.12) the diffusion or white noise limit with averaging, since the rapidly oscillating terms will now be averaged out. Under certain hypotheses we have the following result (cf. Ref. 2). The complex valued process $u_p^\epsilon(\tau)$ converges (weakly) to a diffusion Markov process on $\mathbb{C}^n$, the complex n-dimensional space.

To describe the limiting Markov process we proceed as follows.

We define transport coefficients

$$\begin{aligned}a_{pq,p'q'}^{\cdot,\cdot}&=\Delta(k_p-k_q+k_{p'}-k_{q'})\int_0^\infty\exp[-i(k_p-k_q)\sigma]R_{pq,p'q'}^{\cdot,\cdot}(\sigma)\,d\sigma,\\a_{pq,p'q'}^{*,\cdot}&=\Delta(-k_p+k_q+k_{p'}-k_{q'})\int_0^\infty\exp[+i(k_p-k_q)\sigma]\\&\quad\times R_{pq,p'q'}^{*,\cdot}(\sigma)\,d\sigma,\\a_{pq,p'q'}^{\cdot,*}&=\Delta(-k_p+k_q+k_{p'}-k_{q'})\int_0^\infty\exp[-i(k_p-k_q)\sigma]\\&\quad\times R_{pq,p'q'}^{\cdot,*}(\sigma)\,d\sigma,\\a_{pq,p'q'}^{*,*}&=\Delta(k_p-k_q+k_{p'}-k_{q'})\int_0^\infty\exp[+i(k_p-k_q)\sigma]R_{pq,p'q'}^{*,*}(\sigma)\,d\sigma,\\b_{pq}^{\cdot}&=\Delta(k_p-k_q)\overline{\nu}_{pq},\\b_{pq}^{*}&=\Delta(k_p-k_q)\overline{\nu}_{pq},\quad pq,p'q'=1,2,\ldots,n.\end{aligned} \tag{2.13}$$

Here $\Delta(k)$ is defined to be zero unless $k=0$ when it is equal to 1.

Let $v=v^r+iv^I$, $v^*=v^r-iv^I$, and $\tilde f(v^r,v^I)$ be a smooth function of v^r and v^I. Let

$$f(v,v^*)=\tilde f\left(\frac{v+v^*}{2},\frac{v-v^*}{2i}\right)$$

and define as usual

$$\partial f(v,v^*)=\frac{1}{2}\left(\frac{\partial\tilde f}{\partial v^r}+\frac{\partial\tilde f}{\partial v^I}\right),$$
$$\partial^* f(v,v^*)=\frac{1}{2i}\left(\frac{\partial\tilde f}{\partial v^r}-\frac{\partial\tilde f}{\partial v^I}\right).$$

Let $u=(u_1,u_2,\ldots,u_n)$ be an n-dimensional complex vector and let $f(u,u^*)$ be a smooth function in the sense described above. We denote by ∂_p and ∂_p^* partial derivatives with respect to u_p and u_p^* as above and, using the summation convention, we define the operator $\mathcal{L}$ by

$$\begin{aligned}\mathcal{L}f(u,u^*)&=a_{pq,p'q}^{\cdot,\cdot}u_q\,\partial_p u_{q'}\,\partial_{p'}f(u,u^*)+a_{pq,p'q}^{\cdot,*}u_q\,\partial_p u_{q'}^*\,\partial_{p'}^*f(u,u^*)\\&\quad+a_{pq,p'q}^{*,\cdot}u_q^*\,\partial_p^* u_{q'}\,\partial_{p'}f(u,u^*)+a_{pq,p'q}^{*,*}u_q^*\,\partial_p^* u_{q'}^*\,\partial_{p'}^*f(u,u^*)\\&\quad+b_{pq}u_q\,\partial_p f(u,u^*)+b_{pq}^*u_q^*\,\partial_p^*f(u,u^*)\end{aligned} \tag{2.14}$$

This operator is the infinitesimal generator of the limiting Markov process to which $u^\epsilon(\tau)$ converges that is, $\mathcal{L}_{AD}$, the formal adjoint of $\mathcal{L}$ in (2.14), is the Fokker–Planck operator for $u^0(\tau)$ [the limit of $u^\epsilon(\tau)$]. The transition probability density of $u^0(\tau)$, $P(\tau,u,u^*)$ satisfies the equation

$$\frac{\partial P}{\partial \tau} = \mathcal{L}_{AD} P, \quad P(0,u,u^*) = \delta(u-u_0)\delta(u^*-u_0^*). \tag{2.15}$$

We note that $\mathcal{L}$, and hence $\mathcal{L}_{AD}$, takes real valued functions into real valued functions and that it is a possibly degenerate second order elliptic operator.

From this general convergence result of $u^\epsilon(\tau) \to u^0(\tau)$ one may deduce several interesting consequences.

First, because the stochastic equation (2.12) is linear, we obtain closed equations for moments of each order for $u^0(\tau)$, i.e., as $\epsilon \to 0$. In particular, we obtain closed equations for second and fourth order moments.

Second, and this is important for the applications to Maxwell's equations, the equations for the moments can be obtained independently of the diffusion limit which leads to the Fokker–Planck equation (2.15) and they make sense for partial differential equations (cf. Sec. 4).

Third, the role of the coefficients Δ in (2.13), i.e., of degeneracy or nondegeneracy, is significant. Let us examine it further.

Suppose that (2.7) is nondegenerate, i.e., the numbers k_p along with their sums and differences are distinct and let us consider the equation for the second order moments of $u^0(\tau) = (u_1^0(\tau), \ldots, u_n^0(\tau))$. Let[13]

$$W_p(\tau) = \lim_{\epsilon \to 0} E\{u_p^\epsilon(\tau) u_p^{\epsilon *}(\tau)\} = E\{u_p^0(\tau) u_p^{0*}(\tau)\}, \quad p = 1, 2, \ldots, n. \tag{2.16}$$

From (2.14) and (2.15) we obtain a closed system of equations for the $W_p(\tau)$ alone; no cross moments enter.

$$\frac{dW_p(\tau)}{d\tau} = \sum_{q=1}^n A_{pq} W_q(\tau) + \left(\sum_{q=1}^n B_{pq}\right) W_p(\tau) + C_p W_p(\tau), \quad \tau > 0, \quad W_p(0) = u_{p0} u_{p0}^* \quad p = 1, \ldots, n. \tag{2.17}$$

Here we have introduced the following notation:

$$\begin{aligned} A_{pq} &= \int_0^\infty \exp[i(k_p - k_q)\sigma] R^{*,\cdot}_{pq,pq}(\sigma)\, d\sigma \\ &\quad + \int_0^\infty \exp[-i(k_p - k_q)\sigma] R^{\cdot,*}_{pq,pq}(\sigma)\, d\sigma, \\ B_{pq} &= \int_0^\infty \exp[-i(k_p - k_q)\sigma] R^{\cdot,\cdot}_{pq,pq}(\sigma)\, d\sigma, \\ &\quad + \int_0^\infty \exp[i(k_p - k_q)\sigma] R^{*,*}_{pq,pq}(\sigma)\, d\sigma \\ C_p &= \overline{\nu}_{pp} + \overline{\nu}_{pp}^*. \end{aligned} \tag{2.18}$$

If we specialize further to the case $\mu_{pq}^*(t) = -\mu_{qp}(t)$, $\nu_{pq}(t) \equiv 0$, we obtain a conservative transport equation[10]

$$\frac{dW_p(\tau)}{d\tau} = \sum_{q=1}^n [\tilde{A}_{pq} W_q(\tau) - \tilde{A}_{qp} W_p(\tau)], \quad W_p(0) = u_{p0} u_{p0}^*, \quad \tilde{A}_{pq} = \int_{-\infty}^\infty \cos(k_p - k_q)\sigma R_{pq,pq}(\sigma)\, d\sigma. \tag{2.19}$$

By conservative we mean that $\sum_{p=1}^n W_p(\tau)$ is independent of τ.

The result (2.17), a closed equation for the expectation of the square moduli of $u_p^\epsilon(\tau)$ in the diffusion limit, is clearly a direct consequence of nondegeneracy. In the completely degenerate case, when all the k_p are equal, cross terms enter and (2.17) [or (2.19)] are not valid.

Degenerate problems arise, usually, when the unperturbed problem (2.1) [or (2.7)] has some internal symmetries. It is natural to assume that the stochastic perturbations posses, in a statistical sense, the same symmetries. Is it then true that the quantities (2.16) satisfy equations analogous to (2.17)? The answer is no in general, i.e., cross moment terms cannot be eliminated, and Maxwell's equation (cf. Secs. 3–5) provide an example where degeneracy of the unperturbed equations precludes the validity of equations such as (2.17). However, we have examples where the answer to the above question is affirmative. We present two such examples, beginning with equations (2.20) and (2.37), respectively.

Let $x \in S^2$, the unit sphere in R^3, and consider the stochastic partial differential equation

$$\frac{\partial u(t,x)}{\partial t} = i[\nabla^2 u(t,x) + \epsilon \mu(t,x) u(t,x)], \quad t > 0, \quad u(0,x) = u_0(x), \tag{2.20}$$

for the complex-valued function $u(t,x)$. Here, ∇^2 denotes the Laplace–Beltrami operator on S^2 and $\mu(t,x)$ is a real, stationary, zero mean, random process, almost surely bounded. We assume that its correlation function is rotationally invariant,

$$E\{\mu(t+s,x)\mu(s,x')\} = R(t, x \cdot x'), \tag{2.21}$$

where $x \cdot x'$ is the dot product of the vector x, x' on S^2. Thus, $\mu(t,x)$ is a homogeneous random field on the sphere.

Let $Y_p^l(x)$ denote the normalized spherical harmonics, $p = 0, 1, 2, \cdots$, $-p \le l \le p$, satisfying

$$\int_{S^2} Y_p^l(x) Y_{p'}^{l'*}(x)\, dS(x) = \delta_{pp'}\delta_{ll'}, \tag{2.22}$$

$$\nabla^2 Y_p^l(x) = -p(p+1) Y_p^l(x), \quad -p \le l \le p. \tag{2.23}$$

We expand the solution $u(t,x)$ of (2.20) in spherical harmonics

$$u(t,x) = \sum_{p=0}^\infty \sum_{l=-p}^p v_p^l(t) Y_p^l(x). \tag{2.24}$$

This and (2.20) along with the orthogonality property lead to the equation

$$\frac{dv_p^l(t)}{dt} = -ip(p+1) v_p^l(t) + i\epsilon \sum_{q=0}^\infty \sum_{m=-q}^q \mu_{pq}^{lm}(t) v_q^m(t), \quad t > 0, \quad v_p^l(0) = u_{0,p}^l, \quad p = 0, 1, 2, \cdots, \quad -p \le l \le p, \tag{2.25}$$

where

$$\mu_{pq}^{lm}(t) = \int_{S^2} \mu(t,x) Y_p^{l*}(x) Y_q^m(x)\, dS(x). \tag{2.26}$$

This is a problem in the form (2.1) and it is degenerate since $k_p^l = -p(p+1)$ for $-p \le l \le p$. Let

$$\tau = \epsilon^2 t, \quad u_p^l(\tau;\epsilon) = v_p^l(\tau/\epsilon^2) \exp[ip(p+1)\tau/\epsilon^2]. \tag{2.27}$$

We obtain for $u_p^l(\tau;\epsilon)$ a system of the form (2.12):

$$\frac{du_p^l(\tau;\epsilon)}{d\tau} = \frac{1}{\epsilon} \sum_{q=0}^\infty \sum_{m=-q}^q \exp[ip(p+1)\tau/\epsilon^2] i\mu_{pq}^{lm}(\tau/\epsilon^2) \times \exp[-iq(q+1)\tau/\epsilon^2] u_q^m(\tau;\epsilon), \quad \tau > 0, \quad u_p^l(0;\epsilon) = u_{0,p}^l, \quad p = 0, 1, 2, \cdots, \quad -p \le l \le p. \tag{2.28}$$

The system (2.28) is infinite dimensional while the results quoted above have been shown only for finite dimensional systems. Since we are interested here

primarily in the algebraic calculations that lead to transport equations of the form (2.19), we shall not pause to provide the necessary justification.

Let us expand the covariance of μ, defined by (2.21), is spherical harmonics:

$$R(t,x\cdot x') = \sum_{r=0}^{\infty} R_r(t)\frac{2r+1}{2\pi} P_r(x\cdot x') = \sum_{r=0}^{\infty}\sum_{s=-r}^{r} R_r(t)Y_r^s(x)Y_r^{s*}(x'). \tag{2.29}$$

Here we have employed the addition theorem for spherical harmonics. We next formally compute the diffusion operator (2.14) for the present problem. We find that

$$\mathcal{L} = \sum_{\substack{pq_s p'q' \\ lm_s l'm'}} (a^{\cdot,\cdot\, lm_s l'm'}_{pq_s p'q'} u_q^m \partial_p^l u_{q'}^{m'} \partial_{p'}^{l'} + a^{*,\cdot\, lm_s l'm'}_{pq_s p'q'} u_q^{*m} \partial_p^{*l} u_{q'}^{*m'} \partial_{p'}^{*l'} + a^{\cdot,*\, lm_s l'm'}_{pq_s p'q'} u_q^{*m} \partial_q^{*l} u_{q'}^{m'} \partial_{p'}^{l'} + a^{\cdot,*\, lm_s l'm'}_{pq_s p'q'} u_q^{m} \partial_p^{l} u_{q'}^{*m'} \partial_{p'}^{*l'}), \tag{2.30}$$

where the coefficients a are given by the following formulas

$$a^{\cdot,\cdot\, lm_s l'm'}_{pq_s p'q'} = -\Delta(p(p+1)-q(q+1)+p'(p'+1)-q'(q'+1)) \times \sum_{r=0}^{\infty} \int_0^{\infty} \exp[i(p(p+1)-q(q+1))\sigma] R_r(\sigma)\,d\sigma \times \sum_{S=-r}^{r} \Big(\int Y_r^s Y_p^l Y_q^{m*}\Big)\Big(\int Y_r^{s*} Y_{p'}^{l'*} Y_{q'}^{m'}\Big),$$

$$a^{*,\cdot\, lm_s l'm'}_{pq_s p'q'} = \Delta(-p(p+1)+q(q+1)+p'(p+1)-q'(q'+1)) \times \sum_{r=0}^{\infty} \int_0^{\infty} \exp[-i(p(p+1)-q(q+1))\sigma] R_r(\sigma)\,d\sigma \times \sum_{S=-r}^{r} \Big(\int Y_r^s Y_p^l Y_q^{m*}\Big)\Big(\int Y_r^{s*} Y_{p'}^{l'*} Y_{q'}^{m'}\Big),$$

$$a^{\cdot,*\, lm_s l'm'}_{pq_s p'q'} = \Delta(-p(p+1)+q(q+1)+p'(p'+1)-q'(q'+1)) \times \sum_{r=0}^{\infty} \int_0^{\infty} \exp[i(p(p+1)-q(q+1))\sigma] R_r(\sigma)\,d\sigma \times \sum_{S=-r}^{r} \Big(\int Y_r^s Y_p^l Y_q^{m*}\Big)\Big(\int Y_r^{s*} Y_{p'}^{l'} Y_{q'}^{m'*}\Big),$$

$$a^{*,*\, lm_s l'm'}_{pq_s p'q'} = -\Delta(p(p+1)-q(q+1)+p'(p'+1)-q'(q'+1)) \times \sum_{r=0}^{\infty} \int_0^{\infty} \exp[-i(p(p+1)-q(q+1))\sigma] R_r(\sigma)\,d\sigma \times \sum_{S=-r}^{r} \Big(\int Y_r^s Y_p^l Y_q^{m*}\Big)\Big(\int Y_r^{s*} Y_{p'}^{l'} Y_{q'}^{m'*}\Big). \tag{2.31}$$

Here we have used again the function Δ introduced in (2.13), which has value 1 or zero according as the argument vanishes or not, and integrals of the form

$$\int_{S^2} Y_r^s(x) Y_p^l(x) Y_q^{m*}(x)\,dS(x), \tag{2.32}$$

which can be expressed in terms of Clebsh–Gordan coefficients. It is sufficient for our purposes here to note that the integral (2.32) is zero unless

$$s+l-m=0 \quad \text{and} \quad |p-q| \le r \le p+q, \tag{2.33}$$

which are the selection rules.

As with (2.16) and (2.17) we can use (2.30) to obtain closed equations for

$$W_\gamma^\nu(\tau) = \lim_{\epsilon\to 0} E\{u_\gamma^\nu(\tau;\epsilon)u_\gamma^{*\nu}(\tau;\epsilon)\}, \quad \tau \ge 0, \quad \gamma = 0,1,2,\cdots, \quad -\gamma \le \nu \le \gamma. \tag{2.34}$$

Because of the form of the coefficients (2.31) and the selection rules (2.33), a direct consequence of the rotational invariance of the covariance of $\mu(t,x)$, we find that the W_γ^ν satisfy the following transport equations:

$$\frac{dW_\gamma^\nu(\tau)}{d\tau} = \sum_{q=0}^{\infty}\sum_{m=-q}^{q} [A_{\gamma q}^{\nu m} W_q^m(\tau) - A_{q\gamma}^{m\nu} W_\gamma^\nu(\tau)], \quad \tau > 0, \quad W_\gamma^\nu(0) = u_{0\gamma}^\nu u_{0\gamma}^{*\nu}, \tag{2.35}$$

where

$$A_{\gamma q}^{\nu m} = \sum_{r=|\gamma-q|}^{\gamma+q} \Big|\int Y_r^{m-\nu} Y_\gamma^\nu Y_q^{*m}\Big|^2 \times 2\int_0^{\infty} \cos(\gamma(\gamma+1)-q(q+1))\sigma R_r(\sigma)\,d\sigma, \tag{2.36}$$

and

$$A_{\gamma q}^{\nu m} \ge 0, \quad A_{\gamma q}^{\nu m} = A_{q\gamma}^{m\nu}, \quad \gamma, q = 0,1,2,\cdots, \quad -\gamma \le \nu \le \gamma, \quad -q \le m \le q.$$

We have shown therefore, at least for the truncated system (2.28), that despite the degeneracy of the unperturbed problem (rotational invariance of ∇^2), statistical homogeneity of the random perturbations restores the validity of the transport equations (2.35) in the diffusion limit.

Another example, quite similar to the one just described, is the following. Let g denote an element of $SU(2)$, the group of 2×2 unitary unimodular matrices, and let D^2 denote the Laplace–Beltrami (Casimir) operator on $SU(2)$. Consider the stochastic differential equation

$$\frac{\partial u(t,g)}{\partial t} = i[D^2 u(t,g) + \epsilon\mu(t,g)u(t,g)], \quad t>0, \quad u(0,g) = u_0(g), \tag{2.37}$$

which is analogous to (2.20).

We proceed now in the same way as above. Let $t_{lm}^p(g)$ denote the matrix elements of the $(2p+1)$-dimensional representation of $SU(2)$, $p = 0, \frac{1}{2}, 1, \frac{3}{2}, \cdots, -p \le l, m \le p$, l, m integers or half-odd integers together with p. We note here the following properties[14,15]:

unitarity $$t_{lm}^p(g^{-1}) = t_{lm}^{p*}(g), \tag{2.38}$$

orthogonality $$\int_{SU(2)} t_{l'm'}^{p'}(g) t_{lm}^{p*}(g)\,dg = (2p+1)^{-1}\delta_{pp'}\delta_{ll'}\delta_{mm'}. \tag{2.39}$$

eigenfunction of D^2 $$D^2 t_{lm}^p(g) = -p(p+1)t_{lm}^p(g), \tag{2.40}$$

addition theorem $$t_{lm}^p(g_1 g_2) = \sum_{\nu=-p}^{p} t_{l\nu}^p(g_1) t_{\nu m}^p(g_2). \tag{2.41}$$

We expand the solution of (2.37) in terms of the representation $t_{lm}^p(g)$

$$u(t,g) = \sum_q \sum_{\lambda,\eta} \sqrt{2q+1}\, u_q^{\lambda\eta}(t) t_{\lambda\eta}^q(g), \tag{2.42}$$

so, on using (2.38), we obtain

$$\frac{du_p^{lm}(t)}{dt} = -ip(p+1)u_p^{lm}(t) + i\epsilon\sum_q \sum_{\lambda\eta} \mu_{pq}^{lm_s\lambda\eta}(t) u_q^{\lambda\eta}(t), \quad u_q^{\lambda\eta}(0) = u_{0,q}^{\lambda\eta}. \tag{2.43}$$

Here and in (2.42) the sum is over $q = 0, \frac{1}{2}, 1, \frac{3}{2}, \cdots$ and

$-q \leq \lambda, \eta \leq q$ with integer increments. The coefficients $\mu_{pq}^{lm,\lambda\eta}(t)$ are defined by

$$\mu_{pq}^{lm,\lambda\eta}(t) = \sqrt{(2p+1)(2q+1)} \int_{SU(2)} t_p^{lm*}(g) t_q^{\lambda\eta}(g) \mu(t,g)\, dg. \tag{2.44}$$

Passing to the interaction representation, as in (2.28), we let

$$\tau = \epsilon^2 t, \quad u_p^{lm}(\tau;\epsilon) = v_p^{lm}(\tau/\epsilon^2) \exp[ip(p+1)\tau/\epsilon^2] \tag{2.45}$$

and obtain

$$\frac{du_p^{lm}(\tau;\epsilon)}{d\tau} = \frac{1}{\epsilon} \sum_q \sum_{\lambda\eta} \exp[ip(p+1)\tau/\epsilon^2] i\mu_{pq}^{lm,\lambda\eta}(\tau/\epsilon^2) \times \exp[-iq(q+1)\tau/\epsilon^2] u_q^{\lambda\eta}(\tau;\epsilon),$$

$$\tau > 0, \quad u_p^{lm}(0;\epsilon) = u_{0,p}^{lm}. \tag{2.46}$$

Before evaluating the diffusion limit we must specify the properties of the random function $\mu(t,g)$. It has mean zero, as above, and covariance

$$E\{\mu(t+s,g_1)\mu(s,g_2)\} = R(t,g_1,g_2). \tag{2.47}$$

We assume that R has the following invariance properties:

$$R(t,g_1,g_2) = R(t,e,g_1^{-1}g_2) = R(t,e,g_2g_1^{-1}), \tag{2.48}$$

where e denotes the identity element of $SU(2)$. Thus $R(t,e,g)$ is a function which is constant on conjugacy classes

$$R(t,e,g_2) = R(t,e,g_1g_2g_1^{-1}). \tag{2.49}$$

It has therefore the expansion[15]

$$R(t,g_1,g_2) = R(t,e,g_1^{-1}g_2) = \sum_{r=0}^{\infty} R_r(t) t_{00}^r(g_1^{-1}g_2) = \sum_{r=0}^{\infty} \sum_{\nu=-r}^{r} R_r(t) t_{0\nu}^r(g_1^{-1}) t_{\nu 0}^r(g_2) = \sum_{r=0}^{\infty} \sum_{\nu=-r}^{r} R_r(t) t_{\nu 0}^{r*}(g_1) t_{\nu 0}^r(g_2). \tag{2.50}$$

Here we have used (2.38) and (2.41).

The limiting diffusion operator, analogous to (2.30), can be obtained just as before but the new coefficients a have 12 instead of 8 indices. We shall not write them explicitly. We proceed directly to the transport equations corresponding to (2.35) by employing the selection rules for the Clebsh–Gordan coefficients.[15]

Let

$$W_\gamma^{\nu\lambda}(\tau) = \lim_{\epsilon \to 0} E\{u_\gamma^{\nu\lambda}(\tau;\epsilon) u_\gamma^{\nu\lambda *}(\tau;\epsilon)\}, \quad \tau \geq 0,$$

$$\gamma = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, \cdots, \quad -\gamma \leq \nu, \lambda \leq \gamma. \tag{2.51}$$

Then, $W_\gamma^{\nu\lambda}(\tau)$ satisfy the following transport equations:

$$\frac{dW_\gamma^{\nu\lambda}(\tau)}{d\tau} = \sum_q \sum_{ml} [A_{\gamma q}^{\nu\lambda, ml} W_q^{ml}(\tau) - A_{q\gamma}^{ml,\nu\lambda} W_\gamma^{\nu\lambda}(\tau)],$$

$$\tau > 0, \quad W_\gamma^{\nu\lambda}(0) = u_{0,\gamma}^{\nu\lambda} u_{0,\gamma}^{\nu\lambda *}, \tag{2.52}$$

where

$$A_{\gamma q}^{\nu\lambda, ml} = \sum_{r=|\gamma-q|}^{\gamma+q} (2p+1)(2q+1) \left| \int t_{m-\nu, n-\lambda}^r(g) t_{mn}^r(g) t_{\nu\lambda}^{q*}(g)\, dg \right|^2 \times 2 \int_0^\infty \cos(\gamma(\gamma+1) - q(q+1))\sigma R_r(\sigma)\, d\sigma, \tag{2.53}$$

$$A_{\gamma q}^{\nu\lambda, ml} \geq 0, \quad A_{\gamma q}^{\nu\lambda ml} = A_{q\gamma}^{ml\nu\lambda},$$

$$\gamma, q = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, \cdots, \quad -\gamma \leq \nu, \lambda \leq \gamma, \quad -q \leq m, l \leq q.$$

In (2.53) the summation is over integers or half-odd integers depending on whether $|\gamma - q|$ is an integer or half an odd integer.

3. PRELIMINARY TRANSFORMATIONS OF MAXWELL'S EQUATIONS IN A STATISTICALLY HOMOGENEOUS MEDIUM

In this section we shall treat Maxwell's equations in a nonconducting medium whose dielectric constant[16] ϵ and magnetic permeability μ are random positive definite symmetric tensors. The mean values (ensemble averages) of ϵ and μ will be constant multiples of the unit tensor I so that

$$E\{\epsilon(\mathbf{x},t)\} = \bar{\epsilon} I, \quad E\{\mu(\mathbf{x},t)\} = \bar{\mu} I, \tag{3.1}$$

$\bar{\epsilon}, \bar{\mu}$ are constants. The deviations of ϵ and μ from their means are small stationary tensor-valued random processes in space and time which are of order γ,[16] where γ is a small parameter. The exact nature of these random processes will be specified later.

Maxwell's equations in a nonconducting medium are

$$\frac{1}{c}\partial_t \mathbf{D} = \nabla \wedge \mathbf{H}, \quad \frac{1}{c}\partial_t \mathbf{B} = -\nabla \wedge \mathbf{E}, \tag{3.2}$$

$$\mathbf{D} = \epsilon \mathbf{E}, \quad \mathbf{B} = \mu \mathbf{H}. \tag{3.3}$$

From (3.2) we may deduce that the further equations

$$\nabla \cdot \mathbf{D} = 0, \quad \nabla \cdot \mathbf{B} = 0, \tag{3.4}$$

hold for all times if they hold initially. Here $\mathbf{E}$, $\mathbf{H}$ are the electric displacement and magnetic induction, and c is the speed of light *in vacuo*.

When ϵ, μ are independent of t (3.2), (3.3) imply

$$\frac{1}{c}\epsilon^{1/2}\partial_t(\epsilon^{1/2}\mathbf{E}) = \nabla \wedge \mathbf{H}, \quad \frac{1}{c}\mu^{1/2}\partial_t(\mu^{1/2}\mathbf{H}) = -\nabla \wedge \mathbf{E}, \tag{3.5}$$

where $\epsilon^{1/2}$, $\mu^{1/2}$ are the positive square roots of ϵ, μ. On making the substitution

$$\tilde{\mathbf{E}} = \epsilon^{1/2}\mathbf{E}, \quad \tilde{\mathbf{H}} = \epsilon^{1/2}\mathbf{H}, \tag{3.6}$$

Eqs. (3.5) become

$$\partial_t \tilde{\mathbf{E}} = c\epsilon^{-1/2}\nabla \wedge (\mu^{-1/2}\tilde{\mathbf{H}}), \quad \partial_t \tilde{\mathbf{H}} = -c\mu^{-1/2}\nabla \wedge (\epsilon^{-1/2}\tilde{\mathbf{E}}). \tag{3.7}$$

In what follows we shall take (3.7) to be true even when ϵ, μ do depend upon t since we are primarily interested in the case where the time derivatives of ϵ, μ are small, and (3.7) has the convenient property of energy conservation. Thus we may define the energy density $\mathcal{E}$ by

$$\mathcal{E} = \mathbf{E} \cdot \mathbf{D} + \mathbf{H} \cdot \mathbf{B}, \tag{3.8}$$

which may also be written

$$\mathcal{E} = \tilde{\mathbf{E}}^2 + \tilde{\mathbf{H}}^2. \tag{3.9}$$

Energy conservation follows from (3.5) since $\partial_t \mathcal{E}$ is a divergence

$$\partial_t \mathcal{E} = 2c\nabla \cdot (\mathbf{H} \wedge \mathbf{E}). \tag{3.10}$$

Let us define the matrices A^j by

$$(A^j)_{ik} = \epsilon_{ijk}, \tag{3.11}$$

where ϵ_{ijk} is the alternating symbol, and the matrix $A(\mathbf{k})$, $\mathbf{k} = (k_1, k_2, k_3)$, by

$$A(\mathbf{k}) = A^j k_j, \tag{3.12}$$

with summation implied. Then, if x_1, x_2, x_3 are Cartesian coordinates, $\partial_j = \partial/\partial x_j$, $\partial = (\partial_1, \partial_2, \partial_3)$, we may write (3.7) in the form

$$\begin{aligned} \partial_t \tilde{\mathbf{E}} &= c\epsilon^{-1/2} A(\partial)(\mu^{-1/2}\tilde{\mathbf{H}}), \\ \partial_t \tilde{\mathbf{H}} &= -c\mu^{-1/2} A(\partial)(\epsilon^{-1/2}\tilde{\mathbf{E}}), \end{aligned} \tag{3.13}$$

which, on differentiating the products, becomes

$$\begin{aligned} \partial_t \tilde{\mathbf{E}} &= c\epsilon^{-1/2} A^j \mu^{-1/2} \partial_j \tilde{\mathbf{H}} + c\epsilon^{-1/2} A^j (\partial_j \mu^{-1/2})\tilde{\mathbf{H}}, \\ \partial_t \tilde{\mathbf{H}} &= -c\mu^{-1/2} A^j \epsilon^{-1/2} \partial_j \tilde{\mathbf{E}} - c\mu^{-1/2} A^j (\partial_j \epsilon^{-1/2})\tilde{\mathbf{E}}. \end{aligned} \tag{3.14}$$

Let us now be more explicit about the dependence of the random processes ϵ, μ upon $\mathbf{x}, t$ and the small parameter γ. With no loss in generality and to avoid very lengthy calculations we shall assume that

$$\begin{aligned} \epsilon^{-1/2}(\mathbf{x}, t) &= \epsilon^{(0)} I + \gamma \epsilon^{(1)}(\mathbf{x}, t), \\ \mu^{-1/2}(\mathbf{x}, t) &\equiv I, \end{aligned} \tag{3.15}$$

where $\epsilon^{(0)}$ is a constant scalar, $\epsilon^{(1)}$ is a random symmetric tensor valued process with zero mean and stationary in x, t. Thus, for instance, the correlations

$$E\{\epsilon^{(1)}_{uv}(\mathbf{x}, t)\epsilon^{(1)}_{u'v'}(\mathbf{x}', t')\} \tag{3.16}$$

depend upon $\mathbf{x}, t, \mathbf{x}', t'$ only through $\mathbf{x} - \mathbf{x}'$, $t - t'$. On substituting (3.15) in (3.14), we obtain

$$\begin{aligned} \frac{1}{c}\partial_t \tilde{\mathbf{E}} &= \epsilon^{(0)} A^j \partial_j \tilde{\mathbf{H}} + \gamma \epsilon^{(1)} A^j \partial_j \tilde{\mathbf{H}} \\ \frac{1}{c}\partial_t \tilde{\mathbf{H}} &= -A^j \partial_j \tilde{\mathbf{E}} - \gamma A^j \epsilon^{(1)} \partial_j \tilde{\mathbf{E}} - A^j (\partial_j \epsilon^{(1)})\tilde{\mathbf{E}}. \end{aligned} \tag{3.17}$$

We shall next Fourier transform (3.17) with respect to $\mathbf{x}$. Denote by $\hat{\mathbf{E}}$, $\hat{\mathbf{H}}$, $\hat{\epsilon}^{(1)}$, the Fourier transforms of $\tilde{\mathbf{E}}$, $\tilde{\mathbf{H}}$ and $\epsilon^{(1)}$ with $\mathbf{k}$ the transform variable. Then

$$\begin{aligned} \frac{1}{c}\partial_t \hat{\mathbf{E}}(\mathbf{k}, t) &= -ik_j \epsilon^{(0)} A^j \hat{\mathbf{H}}(\mathbf{k}, t) - i\gamma \int J(\mathbf{k}, \mathbf{s}, t)\hat{\mathbf{H}}(\mathbf{s}, t)\, d\mathbf{s}, \\ \frac{1}{c}\partial_t \hat{\mathbf{H}}(\mathbf{k}, t) &= ik_j A^j \epsilon^{(0)} \hat{\mathbf{E}}(\mathbf{k}, t) + i\gamma \int K(\mathbf{k}, \mathbf{s}, t)\hat{\mathbf{E}}(\mathbf{s}, t)\, d\mathbf{s}, \end{aligned} \tag{3.18}$$

where

$$\begin{aligned} J(\mathbf{k}, \mathbf{s}, t) &= \hat{\epsilon}^{(1)}(\mathbf{k} - \mathbf{s}, t) A^j s_j, \\ K(\mathbf{k}, \mathbf{s}, t) &= A^j k_j \hat{\epsilon}^{(1)}(\mathbf{k} - \mathbf{s}, t). \end{aligned} \tag{3.19}$$

The integrals in (3.18) are over all of R^3.

In order to use the perturbation formalism of the next section, which is formally identical to the one employed in Sec. 2, we must eliminate the terms $O(1)$ as $\gamma \to 0$, on the right side of (3.18). In preparation we first diagonalize the $O(1)$ term which may be written

$$i\epsilon^{(0)} \begin{pmatrix} 0 & -A(\mathbf{k}) \\ A(\mathbf{k}) & 0 \end{pmatrix} \begin{pmatrix} \hat{\mathbf{E}} \\ \hat{\mathbf{H}} \end{pmatrix}. \tag{3.20}$$

The 6×6 block matrix in (3.20) is Hermitian [since $A(\mathbf{k})$ is skew] with eigenvalues $|\mathbf{k}|, |\mathbf{k}|, -|\mathbf{k}|, -|\mathbf{k}|, 0, 0$. The matrix of normalized eigenvectors may be taken to be

$$\tilde{\mathbf{T}}(\mathbf{k}) = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbf{i}(\mathbf{k}) & -\mathbf{j}(\mathbf{k}) & \mathbf{i}(\mathbf{k}) & \mathbf{j}(\mathbf{k}) & \sqrt{2}\hat{\mathbf{k}} & 0 \\ \mathbf{j}(\mathbf{k}) & \mathbf{i}(\mathbf{k}) & -\mathbf{j}(\mathbf{k}) & \mathbf{i}(\mathbf{k}) & 0 & \sqrt{2}\hat{\mathbf{k}} \end{pmatrix} \tag{3.21}$$

where $\hat{\mathbf{k}}$ is the unit vector in the direction $\mathbf{k}$ and $\mathbf{i}(\mathbf{k})$, $\mathbf{j}(\mathbf{k})$, $\hat{\mathbf{k}}$ form a right-handed, orthonormal triple of column vectors in R^3. The dependence of $\mathbf{i}$ and $\mathbf{j}$ on $\mathbf{k}$ will sometimes not be shown explicitly in the sequel.

On writing

$$\Lambda = \mathrm{diag}(1, 1, -1, -1, 0, 0) \tag{3.22}$$

and setting

$$\begin{pmatrix} \hat{\mathbf{E}}(\mathbf{k}, t) \\ \hat{\mathbf{H}}(\mathbf{k}, t) \end{pmatrix} = \tilde{\mathbf{T}}(\mathbf{k}) u(\mathbf{k}, t), \tag{3.23}$$

(3.18) becomes

$$\begin{aligned} \frac{1}{c}\partial_t u(\mathbf{k}, t) &= i\epsilon^{(0)} |\mathbf{k}| \Lambda u(\mathbf{k}, t) \\ &\quad + \gamma \int \tilde{T}^T(\mathbf{k}) K(\mathbf{k}, \mathbf{s}, t) \tilde{T}(\mathbf{s}) u(\mathbf{s}, t)\, d\mathbf{s} \end{aligned} \tag{3.24}$$

where

$$K(\mathbf{k}, \mathbf{s}, t) = \begin{pmatrix} 0 & -iJ(\mathbf{k}, \mathbf{s}, t) \\ iK(\mathbf{k}, \mathbf{s}, t) & 0 \end{pmatrix} \tag{3.25}$$

The $O(1)$ term in (3.24) may be eliminated by the transformation

$$u = \exp(iv|k|\Lambda t)\tilde{w} \tag{3.26}$$

where we have defined

$$v \equiv c\epsilon^{(0)} \tag{3.27}$$

as the speed of wave propagation when $\gamma = 0$. Denoting by $\hat{\epsilon}^{(1)}$ again the quantity $c\hat{\epsilon}^{(1)}$, we have

$$\begin{aligned} \partial_t \tilde{w}(\mathbf{k}, t) &= \int \exp(-iv|k|\Lambda t)\tilde{T}(\mathbf{k}) K(\mathbf{k}, \mathbf{s}, t)\tilde{T}(\mathbf{s}) \\ &\quad \times \exp(iv|s|\Lambda t)\tilde{w}(\mathbf{s}, t)\, d\mathbf{s}. \end{aligned} \tag{3.28}$$

Equation (3.28) is a 6×6 system. We shall reduce it first to a 4×4 system and then further to a 2×2 system. Here we have used the tilde on w so that at a later stage we may use plain w for a different but related quantity.

Consider the divergence equations (3.4). These imply that the components of $\hat{E}$ and $\hat{H}$ along $\mathbf{k}$ do not contribute to the zero-order nor to the first-order terms; they contribute only $O(\gamma^2)$ terms. This contribution has coefficients with mean zero and, according to the perturbation analysis of the next section, it makes no contribution in the asymptotic limit we are seeking.

Thus, in (3.28) the last two components of

$$u(\mathbf{k}, t) = \tilde{T}^T(\mathbf{k}) \begin{pmatrix} \hat{\mathbf{E}}(\mathbf{k}, t) \\ \hat{\mathbf{H}}(\mathbf{k}, t) \end{pmatrix}$$

may be dropped and we may treat (3.20) as a 4×4 system

$$\partial_t \bar{w}(\mathbf{k},t) = \gamma \int \exp(-iv|\mathbf{k}|\bar{\Lambda}t)\bar{T}^T(\mathbf{k})\mathcal{K}(\mathbf{k},\mathbf{s},t)\bar{T}(\mathbf{s}) \times \exp(iv|\mathbf{s}|\bar{\Lambda}t)\bar{w}(\mathbf{s},t)\,d\mathbf{s}, \tag{3.29}$$

where

$$\bar{w} = \begin{pmatrix} 1&0&0&0&0&0\\ 0&1&0&0&0&0\\ 0&0&1&0&0&0\\ 0&0&0&1&0&0 \end{pmatrix} \tilde{w}, \tag{3.30}$$

$$\bar{T}(\mathbf{k}) = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbf{i}(\mathbf{k}) & -\mathbf{j}(\mathbf{k}) & \mathbf{i}(\mathbf{k}) & \mathbf{j}(\mathbf{k}) \\ \mathbf{j}(\mathbf{k}) & \mathbf{i}(\mathbf{k}) & -\mathbf{j}(\mathbf{k}) & \mathbf{i}(\mathbf{k}) \end{pmatrix}, \tag{3.31}$$

and

$$\bar{\Lambda} = \mathrm{diag}(1,1,-1,-1). \tag{3.32}$$

Our next step is to make use of the fact that it is not necessary to distinguish between

$$\begin{pmatrix} \bar{w}_1(\mathbf{k}) \\ \bar{w}_2(\mathbf{k}) \end{pmatrix} \text{ and } \begin{pmatrix} \bar{w}_3(-\mathbf{k}) \\ \bar{w}_4(-\mathbf{k}) \end{pmatrix}$$

which both represent waves travelling in the $+\mathbf{k}$ direction when $\hat{\mathbf{E}}$ and $\hat{\mathbf{H}}$ are multiplied by $\exp(-i\mathbf{k}\cdot x)$ for the synthesis of the inverse Fourier transform. In fact, Eqs. (3.18) have the property that if

$$\begin{pmatrix} \hat{\mathbf{E}}^*(\mathbf{k}) \\ \hat{\mathbf{H}}^*(\mathbf{k}) \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{E}}(-\mathbf{k}) \\ \hat{\mathbf{H}}(-\mathbf{k}) \end{pmatrix} \tag{3.33}$$

is true at $t=0$, it remains true for all time. But (3.33) is true initially since we assume $\tilde{\mathbf{E}}$, $\tilde{\mathbf{H}}$ are real fields. This symmetry is reflected in (3.29) by the fact that

$$\begin{pmatrix} \bar{w}_3(\mathbf{k}) \\ \bar{w}_4(\mathbf{k}) \end{pmatrix} = \begin{pmatrix} \bar{w}_1^*(-\mathbf{k}) \\ \bar{w}_2^*(-\mathbf{k}) \end{pmatrix}, \tag{3.34}$$

so that we may eliminate $\bar{w}_3$ and $\bar{w}_4$ in favor of $\bar{w}_1$ and $\bar{w}_2$.

Let $w(\mathbf{k},t)$ denote the first two components of $\bar{w}(\mathbf{k},t)$ and let us omit the bar so that

$$w(\mathbf{k},t) = \begin{pmatrix} w_1(\mathbf{k},t) \\ w_2(\mathbf{k},t) \end{pmatrix}. \tag{3.35}$$

Let $T(\mathbf{k})$ and $T^\perp(\mathbf{k})$ be given by

$$T(\mathbf{k}) = (\mathbf{i} \quad -\mathbf{j}), \quad T^\perp(\mathbf{k}) = (\mathbf{j} \quad \mathbf{i}) \tag{3.36}$$

where as before $\mathbf{i}, \mathbf{j}, \hat{\mathbf{k}}$ form a right-handed orthonormal triple. Implicit in the notation $T(\mathbf{k})$ is that $\mathbf{i}$ and $\mathbf{j}$ are functions of $\mathbf{k}$ which we now make definite. Let (θ,ϕ) be spherical polar angles relative to some fixed reference frame. Thus, if

$$\hat{\mathbf{k}} = \begin{pmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{pmatrix},$$

then

$$\mathbf{i}(\mathbf{k}) = \begin{pmatrix} \cos\theta\cos\phi \\ \cos\theta\sin\phi \\ -\sin\theta \end{pmatrix}, \quad \mathbf{j}(\mathbf{k}) = \begin{pmatrix} -\sin\phi \\ \cos\phi \\ 0 \end{pmatrix}. \tag{3.37}$$

It is clear that

$$\mathbf{i}(-\mathbf{k}) = \mathbf{i}(\mathbf{k}), \quad \mathbf{j}(-\mathbf{k}) = -\mathbf{j}(\mathbf{k}), \tag{3.38}$$

so that (3.31) and (3.36) imply

$$\bar{T}(\mathbf{k}) = \frac{1}{\sqrt{2}} \begin{pmatrix} T(\mathbf{k}) & T(-\mathbf{k}) \\ T^\perp(\mathbf{k}) & T^\perp(-\mathbf{k}) \end{pmatrix}. \tag{3.39}$$

The matrix $\bar{T}^T(\mathbf{k})\mathcal{K}(\mathbf{k},\mathbf{s},t)\bar{T}(\mathbf{s})$ in (3.29) may be written in partitioned form. Then, using (3.34) the first two rows of (3.39) take the form

$$\begin{aligned} \partial_t w(\mathbf{k},t) = \int \exp[-iv(|\mathbf{k}| - |\mathbf{s}|)t][-iT(\mathbf{k})J(\mathbf{k},\mathbf{s},t)T^\perp(\mathbf{s}) \\ + iT^I(\mathbf{k})K(\mathbf{k},\mathbf{s},t)T(\mathbf{s})]w(\mathbf{s},t)\,d\mathbf{s} \\ + \int \exp[-iv(|\mathbf{k}| + |\mathbf{s}|)t] \\ \times[-iT(\mathbf{k})J(\mathbf{k},\mathbf{s},t)T^\perp(-\mathbf{s})]w^*(-\mathbf{s},t)\,d\mathbf{s} \end{aligned} \tag{3.40}$$

where

$$T^I = (T^\perp)^T. \tag{3.41}$$

From (3.19) it follows that

$$T^T(\mathbf{k})J(\mathbf{k},\mathbf{s},t)T^\perp(\mathbf{s}) = T^T(\mathbf{k})\hat{\epsilon}^{(1)}(\mathbf{k}-\mathbf{s},t)A^j s_j T^\perp(\mathbf{s}).$$

But

$$T^T(\mathbf{k})A^j k_j = |\mathbf{k}|\,T^I(\mathbf{k}), \quad A^j s_j T^\perp(\mathbf{s}) = |\mathbf{s}|\,T(\mathbf{s}).$$

Thus

$$T^T(\mathbf{k})J(\mathbf{k},\mathbf{s},t)T^\perp(\mathbf{s}) = |s|\,T^T(\mathbf{k})\hat{\epsilon}^{(1)}(\mathbf{k}-\mathbf{s},t)T(\mathbf{s}), \tag{3.42}$$

and similarly

$$\begin{aligned} T^I(\mathbf{k})K(\mathbf{k},\mathbf{s},t)T(\mathbf{s}) &= -|\mathbf{k}|\,T^T(\mathbf{k})\hat{\epsilon}^{(1)}(\mathbf{k}-\mathbf{s},t)T(\mathbf{s}), \\ T^T(\mathbf{k})J(\mathbf{k},\mathbf{s},t)T^\perp(-\mathbf{s}) &= |\mathbf{s}|\,T^T(\mathbf{k})\hat{\epsilon}^{(1)}(\mathbf{k}-\mathbf{s},t)T(-\mathbf{s}), \\ T^I(\mathbf{k})K(\mathbf{k},\mathbf{s},t)T(-\mathbf{s}) &= -|\mathbf{k}|\,T^T(\mathbf{k})\hat{\epsilon}^{(1)}(\mathbf{k}-\mathbf{s},t)T(-\mathbf{s}). \end{aligned} \tag{3.43}$$

On using (3.42) and (3.43) in (3.40) and replacing $\mathbf{s}$ by $-\mathbf{s}$ in the second integral, we obtain the equation for $w(\mathbf{k},t)$ in the desired form:

$$\begin{aligned} \partial_t w(\mathbf{k},t) = -i\gamma \int (|\mathbf{k}| + |\mathbf{s}|)\exp[-iv(|\mathbf{k}| - |\mathbf{s}|)t] \\ \times T^T(\mathbf{k})\hat{\epsilon}^{(1)}(\mathbf{k}-\mathbf{s},t)T(\mathbf{s})w(\mathbf{s},t)\,d\mathbf{s} \\ - i\gamma \int (|\mathbf{k}| - |\mathbf{s}|)\exp[-iv(|\mathbf{k}| + |\mathbf{s}|)t] \\ \times T^T(\mathbf{k})\hat{\epsilon}^{(1)}(\mathbf{k}+\mathbf{s},t)T(\mathbf{s})w^*(\mathbf{s},t)\,d\mathbf{s}. \end{aligned} \tag{3.44}$$

Recall that $w(\mathbf{k},t)$ is given by (3.35). Summarizing the above transformations, $w_1(\mathbf{k},t)$ and $w_2(\mathbf{k},t)$ are related to $\hat{\mathbf{E}}(\mathbf{k},t)$, $\hat{\mathbf{H}}(\mathbf{k},t)$ as follows:

$$\begin{aligned} \hat{\mathbf{E}}(\mathbf{k},t) &= \frac{\exp(iv|\mathbf{k}|t)}{\sqrt{2}}(\mathbf{i}(\mathbf{k})w_1(\mathbf{k},t) - \mathbf{j}(\mathbf{k})w_2(\mathbf{k},t)) \\ &\quad + \frac{\exp(-iv|\mathbf{k}|t)}{\sqrt{2}}(\mathbf{i}(-\mathbf{k})w_1^*(-\mathbf{k},t) - \mathbf{j}(-\mathbf{k})w_2^*(-\mathbf{k},t)) \\ \hat{\mathbf{H}}(\mathbf{k},t) &= \frac{\exp(iv|\mathbf{k}|t)}{\sqrt{2}}(\mathbf{j}(\mathbf{k})w_1(\mathbf{k},t) + \mathbf{i}(\mathbf{k})w_2(\mathbf{k},t)) \\ &\quad + \frac{\exp(-iv|\mathbf{k}|t)}{\sqrt{2}}(\mathbf{j}(-\mathbf{k})w_1^*(-\mathbf{k},t) + \mathbf{i}(-\mathbf{k})w_2^*(-\mathbf{k},t)). \end{aligned} \tag{3.45}$$

Let

$$\phi(\mathbf{k},\mathbf{x},t) = v|\mathbf{k}|t - \mathbf{k}\cdot\mathbf{x}, \tag{3.46}$$

$$\mathcal{E}(\mathbf{k},\mathbf{x},t) = \frac{\exp[i\phi(\mathbf{k},\mathbf{x},t)]}{\sqrt{2}}(\mathbf{i}(\mathbf{k})w_1(\mathbf{k},t) - \mathbf{j}(\mathbf{k})w_2(\mathbf{k},t)). \tag{3.47}$$

Then we may rewrite (3.45) in the form

$$\hat{\mathbf{E}}(\mathbf{k},t)\exp(-i\mathbf{k}\cdot\mathbf{x}) = \hat{\mathcal{E}}(\mathbf{k},\mathbf{x},t) + \hat{\mathcal{E}}^*(-\mathbf{k},\mathbf{x},t), \tag{3.48}$$

$$\hat{\mathbf{H}}(\mathbf{k},t)\exp(-i\mathbf{k}\cdot\mathbf{x}) = \hat{\mathbf{k}}\wedge\hat{\mathcal{E}}(\mathbf{k},\mathbf{x},t) - \hat{\mathbf{k}}\wedge\hat{\mathcal{E}}^*(-\mathbf{k},\mathbf{x},t). \tag{3.49}$$

From (3.49) it follows that the magnetic vector $\hat{\mathbf{H}}$ need not be considered independently, so we shall restrict attention to the electric vector. The representation (3.48) of the electric field $\hat{\mathbf{E}}(\mathbf{k},t)\exp(-i\mathbf{k}\cdot\mathbf{x})$ corresponds to a forward propagating wave $\hat{\mathcal{E}}(\mathbf{k},\mathbf{x},t)$ and backward propagating one along $\hat{\mathbf{k}}$. Since in the inverse Fourier transform we integrate $\hat{E}(\mathbf{k},t)\exp(-i\mathbf{k}\cdot\mathbf{x})$ over $\mathbf{k}$, the total wave with wavevector $\mathbf{k}$ is a sum of the forward wave for $\mathbf{k}$ and the backward wave for $-\mathbf{k}$:

$$2\,\mathrm{Re}\,\hat{\mathcal{E}}(\mathbf{k},\mathbf{x},t) = \hat{\mathcal{E}}(\mathbf{k},\mathbf{x},t) + \hat{\mathcal{E}}^*(\mathbf{k},\mathbf{x},t). \tag{3.50}$$

We shall end this section by displaying the relation between the electric wave with vector $\mathbf{k}$ given by (3.50) and the Stokes parameters (Ref. 1, pp. 28–34). Let

$$w_1(\mathbf{k},t) = |w_1(\mathbf{k}.t)|\exp[i\theta_1(\mathbf{k},t)],$$
$$w_2(\mathbf{k},t) = |w_2(\mathbf{k},t)|\exp[i\theta_2(\mathbf{k},t)]. \tag{3.51}$$

With this definition, (3.47) and (3.50) yield

$$2\,\mathrm{Re}\,\hat{\mathcal{E}} = \mathbf{i}(\sqrt{2}\,|w_1|\cos(\phi+\theta_1)) + \mathbf{j}(\sqrt{2}\,|w_2|\cos(\phi+\theta_2+\pi)). \tag{3.52}$$

Define I, Q, U, V as follows:

$$I = 2(|w_1|^2 + |w_2|^2),\quad Q = 2(|w_1|^2 - |w_2|^2),$$
$$U = 4|w_1||w_2|\cos(\theta_1-\theta_2-\pi),$$
$$V = 4|w_1||w_2|\sin(\theta_1-\theta-\pi). \tag{3.53}$$

With this definition we have the identity

$$ww^\dagger = \begin{pmatrix} w_1w_1^* & w_1w_2^* \\ w_2w_1^* & w_2w_2^* \end{pmatrix} = \frac{1}{4}\begin{pmatrix} I+Q & U+iV \\ U-iV & I-Q \end{pmatrix}. \tag{3.54}$$

In (3.53) and (3.54) I, Q, U and V are random functions of $\mathbf{k}$ and t. Equation (3.54) is the desired relationship between the Stokes parameters and the coherence matrix $ww^\dagger$. Relations (3.53) correspond to Chandrasekhar's equations 159 and 160 (Ref. 1, p. 29). Reference should be made to this work for further discussion of their significance.

4. EVOLUTION EQUATION FOR THE COHERENCE MATRIX

In this section we shall apply the perturbation formalism of Sec. 2, to be described shortly for the present problem, to obtain an evolution equation for the expectation of the coherence matrix

$$W(\mathbf{k},t) = w(\mathbf{k},t)w^\dagger(\mathbf{k},t), \tag{4.1}$$

as $\epsilon\to 0$, where $w(\mathbf{k},t)$ satisfies (3.44). Before describing the relevant asymptotic limit we shall derive stochastic equations for the coherence matrix.

We write (3.44) in the form

$$\partial_t w = -i\gamma(K_1 w + K_2 w^*), \tag{4.2}$$

where K_i, $i=1,2$, is the operator which takes one vector function $w(\mathbf{k},t)$ with another $K_i w$ given by

$$(K_i w)(\mathbf{k},t) = \int K_i(\mathbf{k},\mathbf{s},t)w(\mathbf{s},t)\,d\mathbf{s},\quad i=1,2, \tag{4.3}$$

with

$$K_1(\mathbf{k},\mathbf{s},t) = (|\mathbf{k}|+|\mathbf{s}|)\exp[-iv(|\mathbf{k}|-|\mathbf{s}|)t] \times T^T(\mathbf{k})\hat{\epsilon}(\mathbf{k}-\mathbf{s},t)T(\mathbf{s}),$$
$$K_2(\mathbf{k},\mathbf{s},t) = (|\mathbf{k}|-|\mathbf{s}|)\exp[-iv(|\mathbf{k}|+|\mathbf{s}|)t] \times T^T(\mathbf{k})\hat{\epsilon}(\mathbf{k}+\mathbf{s},t)T(\mathbf{s}). \tag{4.4}$$

Here and in the sequel we shall omit the superscript on $\hat{\epsilon}$ since it is not necessary. We note that in (4.2) both w and w^* occur so it is convenient to adjoin to (4.2) its complex conjugate to get

$$\partial_t\begin{pmatrix} w \\ w^* \end{pmatrix} = \gamma\begin{pmatrix} -iK_1 & -iK_2 \\ iK_2^* & iK_1^* \end{pmatrix}\begin{pmatrix} w \\ w^* \end{pmatrix}. \tag{4.5}$$

Now consider the (outer) product

$$\begin{pmatrix} w(\mathbf{k},t) \\ w^*(\mathbf{k},t) \end{pmatrix}(w^\dagger(\mathbf{k}',t)w^T(\mathbf{k}',t)) = \begin{pmatrix} ww^\dagger & ww^T \\ w^*w^T & w^*w^T \end{pmatrix} = \begin{pmatrix} W & V \\ V^* & W^* \end{pmatrix}, \tag{4.6}$$

say, where

$$W^\dagger(\mathbf{k},\mathbf{k}',t) = W(\mathbf{k}',\mathbf{k},t),\quad V^*(\mathbf{k},\mathbf{k}',t) = V(\mathbf{k}',\mathbf{k},t). \tag{4.7}$$

Then, on using (4.5), we see that

$$\partial_t\begin{pmatrix} W & V \\ V^* & W^* \end{pmatrix} = \gamma\begin{pmatrix} -iK_1 & -iK_2 \\ iK_2^* & iK_1^* \end{pmatrix}\begin{pmatrix} W & V \\ V^* & W^* \end{pmatrix} + \gamma\begin{pmatrix} W & V \\ V^* & W^* \end{pmatrix}\begin{pmatrix} iK_1^\dagger & -iK_2^T \\ iK_2^\dagger & -iK_1^T \end{pmatrix}. \tag{4.8}$$

Here, an operator K acting on the left operates on W, V, etc., as functions of k but a K acting on the right operators on the k' dependence. Let

$$\mathcal{W}(\mathbf{k},\mathbf{k}',t) = \begin{pmatrix} W(\mathbf{k},\mathbf{k}',t) & V(\mathbf{k},\mathbf{k}',t) \\ V^*(\mathbf{k},\mathbf{k}',t) & W^*(\mathbf{k},\mathbf{k}',t) \end{pmatrix},$$
$$\tilde{K} = \begin{pmatrix} -iK_1 & -iK_2 \\ iK_2^* & iK_1^* \end{pmatrix}. \tag{4.9}$$

Then (4.8) may be written

$$\partial_t\mathcal{W} = \gamma(\tilde{K}(t)\mathcal{W} + \mathcal{W}\tilde{K}^\dagger(t)) \equiv \gamma K(t)\mathcal{W}, \tag{4.10}$$

say.

We recall that in (4.10) γ is a small parameter and $K(t)$ is a random linear operator such that

$$E\{\tilde{K}(t)\} = 0, \tag{4.11}$$

which follows from the assumption that $\epsilon^{(1)}$ in (3.15) has mean zero. We are therefore in a situation formally similar to (2.1) with $\nu_{pq}\equiv 0$ and the $O(1)$ terms removed by an exponential transformation. Now of course we are dealing with integral operators rather than matrices. The formalism of Sec. 2 can be applied to (4.10) as described in Ref. 17, for example. This perturbation analysis is the Markovian limit of the smoothing approximation.[18,19] A theorem characterizing the asymptotic behavior of solutions to operator equations such as (4.10) is given in Ref. 20 but we shall proceed formally here.

Let $\tau = \gamma^2 t$ and set

$$\mathcal{W}^\gamma(\tau) = \mathcal{W}(\tau/\gamma^2), \tag{4.12}$$

where the arguments $\mathbf{k}, \mathbf{k}'$ are suppressed. Then, according to the procedures described in Refs. 17–20, as $\gamma \to 0$,

$$E\{W^\gamma(\tau)\} \to \overline{W}(\tau)$$

and

$$\partial_\tau \overline{W} = \overline{KK}\,\overline{W} \tag{4.13}$$

where

$$\overline{KK} = \lim_{T\to\infty} \frac{1}{T} \int_0^T \int_0^t E\{K(t)K(t')\}\,dt'\,dt. \tag{4.14}$$

In (4.14) the integrand is the operator composed of $K(t')$ acting first and followed by $K(t)$. Departing somewhat from the framework of Sec. 2, we shall assume that initially $\tilde{\mathbf{E}}(\mathbf{x}, 0)$ and $\tilde{\mathbf{H}}(\mathbf{x}, 0)$ are stationary random fields with mean zero and statistically independent of the random inhomogeneities. This leads to initial data for $\overline{W}$ which contain delta functions. However, since the inhomogeneities are also stationary, the kernel of the operator $\overline{KK}$ will also contain delta functions in a consistent manner so that (4.13) makes sense as we show below.

The explicit calculation of the operator $\overline{KK}$ for (4.10) with K given by (4.9) and K_i, $i = 1, 2$, by (4.3), (4.4) is very lengthy but straightforward. We give a sample calculation in the Appendix. We shall now state the result.

First we note the consequences of stationarity assumed about the random symmetric tensor $\epsilon(\mathbf{x}, t)$ in (3.15) (with the superscript omitted). Let

$$E\{\epsilon_{uv}(\mathbf{x}', t')\epsilon_{u'v'}(\mathbf{x}' - \mathbf{y}, t' - t)\} = R_{uv,\,u'v'}(\mathbf{y}, t),$$
$$u, v, u', v' = 1, 2, 3. \tag{4.15}$$

Then, we have the following:

$$\begin{aligned} &E\{\hat{\epsilon}_{uv}(\mathbf{k}, t')\hat{\epsilon}_{u'v'}(\mathbf{s}, t' - t)\} \\ &= \int\int \exp(i\mathbf{k}\cdot\mathbf{x} + i\mathbf{s}\cdot\mathbf{x}')E\{\epsilon_{uv}(\mathbf{x}, t')\epsilon_{u'v'}(\mathbf{x}', t' - t)\}\,d\mathbf{x}\,d\mathbf{x}' \\ &= \int\int \exp[i(\mathbf{k}+\mathbf{s})\cdot\mathbf{x}]R_{uv,\,u'v'}(\mathbf{y}, t)\exp(-i\mathbf{s}\cdot\mathbf{y})\,d\mathbf{y}\,d\mathbf{x} \\ &= (2\pi)^3\delta(\mathbf{k}+\mathbf{s})\int R_{uv,\,u'v'}(\mathbf{y}, t)\exp(-i\mathbf{s}\cdot\mathbf{y})\,d\mathbf{y} \\ &= (2\pi)^3\delta(\mathbf{k}+\mathbf{s})\hat{R}_{uv,\,u'v'}(-\mathbf{s}, t) \\ &= (2\pi)^3\delta(\mathbf{k}+\mathbf{s})\hat{R}_{uv,\,u'v'}(\mathbf{k}, t). \end{aligned} \tag{4.16}$$

Let S^+ and S^- be defined as follows:

$$\begin{aligned} S^+_{uv,\,u'v'}(\mathbf{k}, \omega) &= (2\pi)^3\int_0^\infty \hat{R}_{uv,\,u'v'}(\mathbf{k}, t)\exp(-iv\omega t)\,dt \\ &= (2\pi)^3\int\int_0^\infty R_{uv,\,u'v'}(\mathbf{y}, t)\exp[i(\mathbf{k}\cdot\mathbf{y} - v\omega t)]\,d\mathbf{y}\,dt, \\ S^-_{uv,\,u'v'}(\mathbf{k}, \omega) &= (2\pi)^3\int_{-\infty}^0 \hat{R}_{uv,\,u'v'}(\mathbf{k}, t)\exp(-iv\omega t)\,dt \\ &= (2\pi)^3\int\int_{-\infty}^0 R_{uv,\,u'v'}(\mathbf{y}, t)\exp[i(\mathbf{k}\cdot\mathbf{y} - v\omega t)]\,d\mathbf{y}\,dt, \\ u, v, u', v' &= 1, 2, 3. \end{aligned} \tag{4.17}$$

Let also

$$S_{uv,\,u'v'} = S^+_{uv,\,u'v'} + S^-_{uv,\,u'v'}, \tag{4.18}$$

which is the space–time power spectral tensor.

As a result of the time averaging in the definition (4.14) of $\overline{KK}$, $\overline{W}$, the top left 2×2 block in $\overline{W}$, decouples from the rest of $\overline{W}$. In addition, if we assume that $\overline{W}(0, \mathbf{k}, \mathbf{k}') = \overline{W}_0(\mathbf{k})\delta(\mathbf{k} - \mathbf{k}')$, which corresponds to stationary initial fields, then this form of $\overline{W}$ is preserved for $\tau > 0$, i.e., $\overline{W}(\tau, \mathbf{k}, \mathbf{k}') = W(\tau, \mathbf{k})\delta(\mathbf{k} - \mathbf{k}')$. We shall continue to denote the "diagonal" terms by $\overline{W}(\tau, \mathbf{k})$ as indicated already. Furthermore, we find that (4.13) leads to the following evolution equation for $\overline{W}(\tau, \mathbf{k}) = \overline{W}_{pp'}(\tau, \mathbf{k})$, $p, p' = 1, 2$ (summation implied):

$$\begin{aligned} \partial_\tau\overline{W}_{pp'}(\tau, \mathbf{k}) &= \int (|\mathbf{k}| + |\mathbf{s}|)^2 T_{up}(\mathbf{k})T_{vq}(\mathbf{s})T_{u'p'}(\mathbf{k})T_{v'q'}(\mathbf{s}) \\ &\quad\times S_{uv,\,u'v'}(\mathbf{k}-\mathbf{s}, |\mathbf{k}| - |\mathbf{s}|)\overline{W}_{qq'}(\tau, \mathbf{s})\,d\mathbf{s} \\ &\quad + \int (|\mathbf{k}| - |\mathbf{s}|)^2 T_{up}(\mathbf{k})T_{vq}(\mathbf{s})T_{u'p'}(\mathbf{k})T_{v'q'}(\mathbf{s}) \\ &\quad\times S_{uv,\,u'v'}(\mathbf{k}+\mathbf{s}, |\mathbf{k}| + |\mathbf{s}|)\overline{W}^*_{qq'}(\tau, \mathbf{s})\,d\mathbf{s} \\ &\quad - \int T_{up}(\mathbf{k})T_{vr}(\sigma)T_{u'r}(\sigma)T_{v'q}(\mathbf{k})[(|\mathbf{k}| + |\sigma|)^2 \\ &\quad\times S^+_{uv,\,u'v'}(\mathbf{k}-\sigma, |\mathbf{k}| - |\sigma|) + (|\mathbf{k}| - |\sigma|)^2 \\ &\quad\times S^+_{uv,\,u'v'}(\mathbf{k}+\sigma, |\mathbf{k}| + |\sigma|)]\,d\sigma\cdot\overline{W}_{qp'}(\tau, \mathbf{k}) \\ &\quad - \int T_{up'}(\mathbf{k})T_{vr}(\sigma)T_{u'r}(\sigma)T_{v'q'}(\mathbf{k})[(|\mathbf{k}| + |\sigma|)^2 \\ &\quad\times S^{+*}_{uv,\,u'v'}(\mathbf{k}-\sigma, |\mathbf{k}| - |\sigma|) + (|\mathbf{k}| - |\sigma|)^2 \\ &\quad\times S^{+*}_{uv,\,u'v'}(\mathbf{k}+\sigma, |\mathbf{k}| + |\sigma|)]\,d\sigma\cdot\overline{W}_{pq'}(\tau, \mathbf{k}). \end{aligned} \tag{4.19}$$

This evolution equation is the main result of this section. In the next section we shall specialize (4.19) and compare it with Chandrasekhar's[1] equations. In the remainder of this section we shall verify that Eq. (4.19) has certain properties which are necessary for a reasonable transport equation. These are:

(a) Under the "reversibility" hypothesis

$$S_{uv,\,u'v'} = S^{+*}_{u'v',\,uv} + S^{+*}_{uv,\,u'v'} = S^*_{uv,\,u'v'} \tag{4.20}$$

which follows if

$$R_{uv,\,u'v'}(\mathbf{y}, t) = R_{uv,\,u'v'}(-\mathbf{y}, -\mathbf{t}), \tag{4.21}$$

i.e., $\epsilon(\mathbf{x}, t)$ is reversible, Eq. (4.19) conserves "total energy":

$$\partial_\tau\int (\overline{W}_{11}(\tau, \mathbf{k}) + \overline{W}_{22}(\tau, \mathbf{k}))\,d\mathbf{k} = 0. \tag{4.22}$$

(b) For each $\mathbf{k}$ the positive definiteness of $\overline{W}(\tau, \mathbf{k})$ is preserved.

Let us show that (4.22) holds. Let us take ∂_τ under the integral sign. Then it is just the left member of (4.19) contracted over p, p' and integrated on $\mathbf{k}$. The right member similarly contracted and integrated is

$$\begin{aligned} &\int\int (|\mathbf{k}| + |\mathbf{s}|)^2 T_{up}(\mathbf{k})T_{vq}(\mathbf{s})T_{up}(\mathbf{k})T_{v'q'}(\mathbf{s}) \\ &\quad\times S_{uv,\,u'v'}(\mathbf{k}-\mathbf{s}, |\mathbf{k}| - |\mathbf{s}|)\overline{W}_{qq}(\tau, \mathbf{s})\,d\mathbf{s}\,d\mathbf{k} \\ &+ \int\int (|\mathbf{k}| - |\mathbf{s}|)^2 T_{up}(\mathbf{k})T_{vq}(\mathbf{s})T_{u'p}(\mathbf{k})T_{v'q'}(\mathbf{s}) \\ &\quad\times S_{uv,\,u'v'}(\mathbf{k}+\mathbf{s}, |\mathbf{k}| + |\mathbf{s}|)\overline{W}^*_{qq'}(\tau, \mathbf{s})\,d\mathbf{s}\,d\mathbf{k} \\ &- \int\int (|\mathbf{k}| + |\mathbf{s}|)^2 T_{up}(\mathbf{k})T_{vr}(\sigma)T_{u'r}(\sigma)T_{v'q}(\mathbf{k}) \\ &\quad\times S^+_{uv,\,u'v'}(\mathbf{k}-\sigma, |\mathbf{k}| - |\sigma|)\overline{W}_{qp}(\tau, \mathbf{k})\,d\sigma\,d\mathbf{k} \\ &- \int\int (|\mathbf{k}| - |\mathbf{s}|)^2 T_{up}(\mathbf{k})T_{vr}(\sigma)T_{u'r}(\sigma)T_{v'q}(\mathbf{k}) \\ &\quad\times S^+_{uv,\,u'v'}(\mathbf{k}+\sigma, |\mathbf{k}| + |\sigma|)\overline{W}_{qp}(\tau, \mathbf{k})\,d\sigma\,d\mathbf{k} \\ &- \int\int (|\mathbf{k}| + |\sigma|)^2 T_{up}(\mathbf{k})T_{vr}(\sigma)T_{u'r}(\sigma)T_{v'q}(\mathbf{k}) \\ &\quad\times S^{+*}_{uv,\,u'v'}(\mathbf{k}-\sigma, |\mathbf{k}| - |\sigma|)\overline{W}_{pq}(\tau, \mathbf{k})\,d\sigma\,d\mathbf{k} \\ &- \int\int (|\mathbf{k}| - |\sigma|)^2 T_{up}(\mathbf{k})T_{vr}(\sigma)T_{u'r}(\sigma)T_{v'q}(\mathbf{k}) \\ &\quad\times S^{+*}_{uv,\,u'v'}(\mathbf{k}+\sigma, |\mathbf{k}| + |\sigma|)\overline{W}_{pq}(\tau, \mathbf{k})\,d\sigma\,d\mathbf{k} \end{aligned} \tag{4.23}$$

We shall show that the third and fifth integrals combine to cancel the first and similarly the fourth and sixth cancel the second.

First $\overline{W}^*_{pq} = \overline{W}_{qp}$ and so we may combine the third and fifth terms to get

$$-\int\int (|\mathbf{k}| + |\sigma|)^2 T_{up}(\mathbf{k}) T_{vr}(\sigma) T_{u'r}(\sigma) T_{v'q}(\mathbf{k}) \times (S^*_{uv,\,u'v'} + S^{**}_{v'u',\,vu})(\mathbf{k}-\sigma, |\mathbf{k}| - |\sigma|) \overline{W}_{pq}(\tau, \mathbf{k})\, d\sigma\, d\mathbf{k}. \quad (4.24)$$

On renaming $\mathbf{k}$ as $\mathbf{s}$, σ as $\mathbf{k}$, r as p, p as q', u as v', v' as v and u' as u, we get

$$-\int\int (|\mathbf{k}| + |\mathbf{s}|)^2 T_{v'q'}(\mathbf{s}) T_{u'p}(\mathbf{k}) T_{up}(\mathbf{k}) T_{vq}(\mathbf{k}) \times (S^*_{v'u'uv} + S^{**}_{vuu'v'})(\mathbf{s}-\mathbf{k}, |\mathbf{s}| - |\mathbf{k}|) \overline{W}_{qq'}(\tau, \mathbf{s})\, d\mathbf{k}\, d\mathbf{s}. \quad (4.25)$$

But S^* is symmetric in its first two subscripts since it inherits this symmetry from the symmetry of $\epsilon(\mathbf{x}, t)$. Thus, in view of (4.20) the cancellation with the first term in (4.23) follows.

Let us next show that (4.19) preserves $\overline{W}(\tau, \mathbf{k})$ as a positive definite hermitian matrix. It is clear using (4.20) that (4.19) preserves hermiticity. Let us rewrite (4.19) in the more convenient form

$$\partial_\tau \overline{W}(\tau, \mathbf{k}) = \int S_1[\mathbf{k}, \mathbf{s}, \overline{W}(\tau, \mathbf{s})]\, d\mathbf{s} + \int S_2[\mathbf{k}, \mathbf{s}, \overline{W}^*(\tau, \mathbf{s})]\, d\mathbf{s} - A(\mathbf{k})\overline{W}(\tau, \mathbf{k}) - \overline{W}(\tau, \mathbf{k})A^*(\mathbf{k}), \quad (4.26)$$

where the operators S_1 and S_2 are linear in the third slot and $A(\mathbf{k})$ is a matrix. Transferring the last two terms to the left in (4.26), we get

$$\partial_\tau Z(\tau, \mathbf{k}) = \exp[-A(\mathbf{k})\tau] \int S_1[\mathbf{k}, \mathbf{s}, \overline{W}(\tau, \mathbf{s})]\, d\mathbf{s} \exp[-A^\dagger(\mathbf{k})\tau] + \exp[-A(\mathbf{k})\tau] \int S_2[\mathbf{k}, \mathbf{s}, \overline{W}^*(\tau, \mathbf{s})]\, d\mathbf{s} \times \exp[-A^\dagger(\mathbf{k})\tau], \quad (4.27)$$

where

$$Z(\tau, \mathbf{k}) = \exp[A(\mathbf{k})\tau] \overline{W}(\tau, \mathbf{k}) \exp[A^\dagger(\mathbf{k})\tau]. \quad (4.28)$$

From (4.28) it follows that $\overline{W}$ is nonnegative if and only if Z is. Moreover, $\partial_\tau Z$ is nonnegative if and only if

$$\int S_1[\mathbf{k}, \mathbf{s}, \overline{W}(\tau, \mathbf{s})]\, d\mathbf{s} + \int S_2[\mathbf{k}, \mathbf{s}, \overline{W}^*(\tau, \mathbf{s})]\, d\mathbf{s} \quad (4.29)$$

is nonnegative. Thus it is enough that (4.29) is nonnegative if $\overline{W}$ is. We proceed to show this next for the second term in (4.29) since the other one follows in the same way.

Let ξ be an arbitrary vector. We shall show that for each $\mathbf{k}$ and $\mathbf{s}$

$$\xi S_2[\mathbf{k}, \mathbf{s}, \overline{W}^*(\tau, \mathbf{s})]\xi^\dagger \geq 0, \quad (4.30)$$

provided that $\overline{W}$ is nonnegative. If $\overline{W}$ is nonnegative $\overline{W}^*$ can be expressed as a sum of two terms each of which is nonnegative

$$\overline{W}^* = \eta\eta^\dagger + \zeta\zeta^\dagger$$

where η and ζ are eigenvectors normalized so that $|\eta|^2$ and $|\zeta|^2$ are the (positive) eigenvalues of $\overline{W}^*$. From the linearity of S_2 in the third slot, it is enough to consider

$$\xi S_2[\mathbf{k}, \mathbf{s}, \eta\eta^\dagger]\xi^\dagger. \quad (4.31)$$

Writing this explicitly from (4.19), we get

$$(|\mathbf{k}| - |\mathbf{s}|)^2 T_{up}(\mathbf{k}) \xi_p T_{vq}(\mathbf{s}) \eta^*_q T_{u'p'}(\mathbf{k}) \xi^*_{p'} T_{v'q'}(\mathbf{s}) \eta_{q'} \times S_{vuu'v'}(\mathbf{k}+\mathbf{s}, |\mathbf{k}| + |\mathbf{s}|). \quad (4.32)$$

Let $T(\mathbf{k})\xi = \alpha$ and $T(\mathbf{s})\eta^* = \beta$. Then to show (4.32) is positive it is enough to show

$$\alpha_u \beta_v \alpha^*_{u'} \beta^*_{v'} S_{uvu'v'}(\mathbf{k}+\mathbf{s}, |\mathbf{k}| + |\mathbf{s}|) \geq 0. \quad (4.33)$$

But S is a power-spectral tensor and therefore by Bochner's theorem (4.33) is true. The proof of preservation of nonnegativity for $\overline{W}$ in (4.19) is complete.

5. THE TIME-HOMOGENEOUS ISOTROPIC CASE

Let us assume that the random inhomogeneous $\epsilon(\mathbf{x}, t)$ in (3.15) do not depend on t and that in (4.15) we have

$$R_{uv,\,u'v'}(\mathbf{y}, t) = \delta_{uv}\delta_{u'v'} R(|\mathbf{y}|). \quad (5.1)$$

From (4.17) and (4.18) it follows that

$$\begin{aligned} S_{uv,\,u'v'}(\mathbf{k}, w) &= (2\pi)^3 \int_{-\infty}^{\infty} R(|\mathbf{y}|) \exp[i(\mathbf{k}\cdot\mathbf{y} - \omega v t)]\, d\mathbf{y}\, dt\, \delta_{uv}\delta_{u'v'} \\ &= (2\pi)^5 \frac{\delta(\omega)}{v} \int_0^\infty \int_0^\pi R(r) \exp(i|\mathbf{k}| r\cos\theta) \times r^2 \sin\theta\, d\theta\, dr\, \delta_{uv}\delta_{u'v'} \\ &= 2\pi\delta_{uv}\delta_{u'v'} \frac{\delta(\omega)}{v} 2\cdot(2\pi)^4 \int_0^\infty \frac{r\sin|\mathbf{k}|r}{|\mathbf{k}|} R(r)\, dr \\ &= 2\pi\delta_{uv}\delta_{u'v'} \frac{\delta(\omega)}{v} \tilde{S}(|k|), \end{aligned} \quad (5.2)$$

say, and similarly

$$S^*_{uv,\,u'v'}(\mathbf{k}, w) = \delta_{uv}\delta_{u'v'} \left(\pi \frac{\delta(\omega)}{v} + \frac{i}{\omega v}\right) \tilde{S}(|k|). \quad (5.3)$$

In (5.3) $1/\omega$ denotes the generalized function corresponding to the principal value and this will be the meaning of the singular integrals in (5.4) below.

Using polar coordinates in $\mathbf{s}$ so that $\mathbf{s} = |\mathbf{s}|\xi$ and $d\mathbf{s} = d\Omega(\hat{\mathbf{s}})|\mathbf{s}|^2 d|\mathbf{s}|$ we rewrite (4.19) for S and S^* given by (5.2) and (5.3). After a few rearrangements we obtain

$$\begin{aligned} \partial_\tau \overline{W}_{pp'}(\tau, \mathbf{k}) = {} & 2\pi \int_{|\mathbf{k}|=|\mathbf{s}|} \frac{4|\mathbf{k}|^4}{v} T_{up}(\hat{\mathbf{k}}) T_{uq}(\hat{\mathbf{s}}) T_{u'p'}(\hat{\mathbf{k}}) T_{u'q'}(\hat{\mathbf{s}}) \\ & \times \tilde{S}(|\mathbf{k}|\,|\hat{\mathbf{k}} - \hat{\mathbf{s}}|) \overline{W}_{qq'}(\tau, |\mathbf{k}|\hat{\mathbf{s}})\, d\Omega(\hat{\mathbf{s}})\, d|\mathbf{s}| \\ & - \pi \frac{4|\mathbf{k}|^4}{v} \int_{|\mathbf{k}|=|\sigma|} T_{up}(\hat{\mathbf{k}}) T_{ur}(\hat{\sigma}) T_{u'r}(\hat{\sigma}) T_{u'q}(\hat{\mathbf{k}}) \tilde{S}(|\mathbf{k}|\,|\hat{\mathbf{k}} - \hat{\sigma}|)\, d\Omega(\hat{\sigma}) \overline{W}_{qp'}(\tau, \mathbf{k}) \\ & + \frac{i}{v} \int_0^\infty \frac{(|\mathbf{k}| + |\sigma|)^2 |\sigma|^2}{|\mathbf{k}| - |\sigma|} \int T_{up}(\hat{\mathbf{k}}) T_{ur}(\hat{\sigma}) T_{u'r}(\hat{\sigma}) T_{u'q}(\hat{\mathbf{k}}) \tilde{S}(|\mathbf{k}|\,|\hat{\mathbf{k}} - \hat{\sigma}|)\, d\Omega(\hat{\sigma})\, d|\sigma|\, \overline{W}_{qp'}(\tau, \mathbf{k}) \end{aligned}$$

$$+\frac{i}{v}\int_0^\infty \frac{(|\mathbf{k}|-|\sigma|)^2|\sigma|^2}{|\mathbf{k}|+|\sigma|}\int T_{up}(\hat{\mathbf{k}})T_{ur}(\hat{\sigma})T_{u'r}(\hat{\sigma})T_{u'q}(\hat{\mathbf{k}})\tilde{S}(|\mathbf{k}|\,|\hat{\mathbf{k}}-\hat{\sigma}|)\,d\Omega(\sigma)\,d|\sigma|\,\overline{W}_{qp'}(\tau,\mathbf{k})$$

$$-\pi\frac{4|\mathbf{k}|^4}{v}\int T_{up'}(\hat{\mathbf{k}})T_{ur}(\hat{\sigma})T_{u'r}(\hat{\sigma})T_{u'q'}(\hat{\mathbf{k}})\tilde{S}(|\mathbf{k}|\,|\hat{\mathbf{k}}-\hat{\sigma}|)\,d\Omega(\hat{\sigma})\,d|\sigma|\,\overline{W}_{pq'}(\tau,\mathbf{k})$$

$$-\frac{i}{v}\int_0^\infty \frac{(|\mathbf{k}|+|\sigma|)^2|\sigma|^2}{|\mathbf{k}|-|\sigma|}\int T_{up'}(\hat{\mathbf{k}})T_{ur}(\hat{\sigma})T_{u'r}(\hat{\sigma})T_{u'q'}(\hat{\mathbf{k}})\tilde{S}(|\mathbf{k}|\,|\hat{\mathbf{k}}-\hat{\sigma}|)\,d\Omega(\hat{\sigma})\,d|\sigma|\,\overline{W}_{pq'}(\tau,\mathbf{k})$$

$$-\frac{i}{v}\int_0^\infty \frac{(|\mathbf{k}|-|\sigma|)^2|\sigma|^2}{|\mathbf{k}|+|\sigma|}\int T_{up'}(\hat{\mathbf{k}})T_{ur}(\hat{\sigma})T_{u'r}(\hat{\sigma})T_{u'q'}(\hat{\mathbf{k}})\tilde{S}(|\mathbf{k}|\,|\hat{\mathbf{k}}-\hat{\sigma}|)\,d\Omega(\sigma)\,d|\sigma|\,\overline{W}_{pq'}(\tau,\mathbf{k}). \tag{5.4}$$

Let us define the 2×2 matrix

$$L=(L_{pq}(\mathbf{k},\mathbf{s}))=(T_{up}(\mathbf{k})T_{uq}(\mathbf{s})). \tag{5.5}$$

We note that

$$L^T(\mathbf{k},\sigma)=L(\sigma,\mathbf{k}). \tag{5.6}$$

Furthermore, recalling that $T(\hat{\mathbf{k}})$ is given by (3.36) and i, j by (3.37), it is easily verified that

$$\int L(\hat{\mathbf{k}},\hat{\sigma})L(\hat{\sigma},\hat{\mathbf{k}})\,d\Omega(\sigma)=\frac{8\pi}{3}I. \tag{5.7}$$

For $|\mathbf{k}|$ fixed assume that $\tilde{S}(\alpha)\equiv\bar{S}$ (constant) say, for $0\le\alpha\le 2|\mathbf{k}|$ and $S(\alpha)=0$ for α large. In this case (5.4) simplifies considerably because the imaginary terms cancel with the help of (5.7). Thus, for this choice of power spectrum $\tilde{S}$ we have

$$\partial_\tau\overline{W}(\tau,\mathbf{k})=\frac{64\pi^2|\mathbf{k}|^4\bar{S}}{3v}\left(\frac{1}{4\pi}\int_{|\mathbf{k}|=|\mathbf{s}|}L(\hat{\mathbf{k}},\hat{\mathbf{s}})\overline{W}(\tau,|\mathbf{k}|\hat{\mathbf{s}})\right.$$
$$\left.\times L^T(\hat{\mathbf{k}},\hat{\mathbf{s}})\,d\Omega(\mathbf{s})-\overline{W}(\tau,\mathbf{k})\right). \tag{5.8}$$

The evolution equation (5.8) is a transport equation for the average coherence matrix $\overline{W}(\tau,\mathbf{k})$ in the asymptotic limit of long times or distances of propagation and weak inhomogeneities. We have assumed throughout statistical spatial homogeneity and so the size of the wavenumber $|\mathbf{k}|$ did not play any role in the asymptotics. If we interpret τ in (5.8) as optical distance, rather than time, then (5.8) coincides (up to normalization) with Chandrasekhar's equation (2.12) (Ref. 1, p. 40) including the $|\mathbf{k}|^4$ dependence of the constant, which appears as the λ^{-4} in Chandrasekhar, originating from Rayleigh's law of scattering. This follows immediately from the relation

$$\overline{W}(\tau,\mathbf{k})=\frac{1}{4}\begin{pmatrix}\bar{I}+\bar{Q} & \bar{U}+i\bar{V}\\ \bar{U}-I\bar{V} & \bar{I}-\bar{Q}\end{pmatrix}, \tag{5.9}$$

corresponding to (3.54) between the average coherence matrix and the mean values of the Stokes parameters $\bar{I}$, $\bar{Q}$, $\bar{U}$, $\bar{V}$, and the definition (5.5) of L.

It is interesting to note that Rayleigh's law of scattering that gave rise to Eq. (2.12) in Ref. 1 corresponds here to assumption (5.1) and the further assumption on $\tilde{S}$ stated below (5.7). If this latter assumption is not satisfied, however, the form of (5.8) changes rather drastically since we have to go back to (5.4). Thus, if the power spectrum $\tilde{S}(\alpha)$ of the inhomogeneities varies significantly in the region $0\le\alpha\le 2|\mathbf{k}|$, Eq. (5.8) is no longer valid.

When the assumption of statistical spatial homogeneity is replaced by (i) local spatial statistical homogeneity, (ii) $|\mathbf{k}|$ large compared to the reciprocal characteristic length of these variations then, we expect that (5.8) should be replaced by a space–time transport equation of the local average coherence matrix $\overline{W}(\tau,\mathbf{k},\mathbf{x})$. Note that for an asymptotic theory of this sort the size of the wavenumber of the primary fields plays an important role, in addition to γ, the size of the inhomogeneities. As we mentioned in the introduction, such a result requires additional considerations and will not be given here.

We note finally that the imaginary terms in (5.4) may be interpreted as representing residual phase retardation effects. This interpretation is motivated by the similar appearance these terms have to the real part of the effective propagation constant in the dispersion theory for the mean fields.[18]

APPENDIX. THE OPERATOR $\overline{KK}$ OF (4.14)

In (4.14) we defined the operator

$$\overline{KK}=\lim_{T\to\infty}\frac{1}{T}\int_0^T\int_0^t E\{K(t)K(t')\}\,dt'\,dt. \tag{A1}$$

Here

$$K(t)W=\tilde{K}(t)W+W\tilde{K}^T(t), \tag{A2}$$

where

$$W(\mathbf{k},\mathbf{k}',t)=\begin{pmatrix}W(\mathbf{k},\mathbf{k}',t) & V(\mathbf{k},\mathbf{k}',t)\\ V^*(\mathbf{k},\mathbf{k}',t) & W^*(\mathbf{k},\mathbf{k}',t)\end{pmatrix}, \tag{A3}$$

$$\tilde{K}=\begin{pmatrix}-iK_1 & -iK_2\\ iK_2^* & iK_1^*\end{pmatrix}, \tag{A4}$$

K_1, K_2 being operators defined on 2×2 matrix functions $W(k)$ by

$$(K_1W)(\mathbf{k})=\int(|\mathbf{k}|+|\mathbf{s}|)\exp[-iv(|\mathbf{k}|-|\mathbf{s}|)t]$$
$$\times T^T(\mathbf{k})\hat{\epsilon}(\mathbf{k}-\mathbf{s},t)T(\mathbf{s})W(\mathbf{s})\,d\mathbf{s},$$
$$(K_2W)(\mathbf{k})=\int(|\mathbf{k}|-|\mathbf{s}|)\exp[-iv(|\mathbf{k}|+|\mathbf{s}|)t]$$
$$\times T^t(\mathbf{k})\hat{\epsilon}(\mathbf{k}+\mathbf{s},t)T(\mathbf{s})W(\mathbf{s})\,d\mathbf{s}. \tag{A5}$$

Using (A2) we have

$$K(t)K(t')W=\tilde{K}(t)\tilde{K}(t')W+W\tilde{K}^T(t')\tilde{K}^\dagger(t)$$
$$+K(t)W\tilde{K}^\dagger(t')+K(t')W\tilde{K}^T(t), \tag{A6}$$

where W is a dummy operand and is regarded as deterministic when taking expectations for (A1).

To illustrate the calculation we shall consider only the contribution of the first term on the right of (A6). From (A4)

$$\tilde{K}(t)\tilde{K}(t') = \begin{pmatrix} -K_1(t)K_1(t')+K_2(t)K_2(t') & -K_1(t)K_2(t')+K_2(t)K_1(t') \\ K_2^*(t)K_1(t')-K_1^*(t)K_2(t') & K_2(t)K_2(t')-K_1(t)K_1(t') \end{pmatrix} \tag{A7}$$

For the purposes of (A1) we need

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T\int_0^t E\{\tilde{K}(t)\tilde{K}(t')\}\,dt'\,dt. \tag{A8}$$

The top left block of this is

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T\int_0^t E\{-K_1(t)K_1(t')+K_2(t)K_2(t')\}\,dt'\,dt. \tag{A9}$$

From (A5) the first term in (A9) is a linear integral operator acting upon 2×2 matrices whose kernel is the matrix function

$$-\lim_{T\to\infty}\frac{1}{T}\int_0^T\int_0^t\int_{R^3}(|\mathbf{k}|+|\sigma|)(|\sigma|+|\mathbf{s}|) \times\exp(-iv[(|\mathbf{k}|-|\sigma|)t+(|\sigma|-|\mathbf{s}|)t']) \times E\{T^T(\mathbf{k})\hat{\epsilon}(\mathbf{k}-\sigma,t)T(\sigma)T^T(\sigma)\hat{\epsilon}(\sigma-\mathbf{s},t')T(\mathbf{s})\}\,d\sigma\,dt'\,dt. \tag{A10}$$

The pq entry of this may be written

$$-\lim_{T\to\infty}\frac{1}{T}\int_0^T\int_0^T\int_{R^3}(|\mathbf{k}|+|\sigma|)(|\sigma|+|\mathbf{s}|) \times\exp(-iv[(|\mathbf{k}|-|\sigma|)t+(|\sigma|-|\mathbf{s}|)(t-\tau)]) \times T_{up}(\mathbf{k})T_{ur}(\sigma)T_{u'r}(\sigma)T_{v'q}(\mathbf{s}) \times E\{\hat{\epsilon}_{uv}(\mathbf{k}-\sigma,t)\hat{\epsilon}_{u'v'}(\sigma-\mathbf{s},t-\tau)\}\,d\sigma\,dt\,d\tau, \tag{A11}$$

where we have changed variable of integration from t' to $\tau=t-t'$ and inverted the order of integration. From (4.16) we may write this as

$$-(2\pi)^3\int(|\mathbf{k}|+|\sigma|)^2\delta(\mathbf{k}+\mathbf{s})T_{up}(\mathbf{k})T_{vr}(\sigma)T_{u'r}(\sigma)T_{v'q}(\mathbf{s}) \times\lim_{T\to\infty}\frac{1}{T}\int_0^T\int_\tau^T\hat{R}_{uv,u'v'}(\mathbf{k}-\sigma,\tau)\exp[iv(|\mathbf{k}|-|\sigma|)\tau]\,dt\,d\tau\,d\sigma$$
$$=\frac{1}{\epsilon}(2\pi)^3\delta(\mathbf{k}-\mathbf{s})\int(|\mathbf{k}|+|\sigma|)^2T_{up}(\mathbf{k})T_{vr}(\sigma)T_{u'r}(\sigma)T_{v'q}(\mathbf{s}) \times S^+_{uv,u'v'}(\mathbf{k}-\sigma,|\mathbf{k}|-|\sigma|)\,d\sigma. \tag{A12}$$

This result contributes the fourth term on the right in (4.19).

This is the contribution of $-K_1(t)K(t')$ in (A7). The contributions of the other terms may be calculated in a similar way. In like manner every part of the various terms in (A6) may be calculated.

It is found that the operator $\overline{KK}$ of (A1) is reducible in that the blocks containing W, W^* in (A3) are invariant under the action $\overline{KK}$, and so are the blocks containing V, V^*. [It is found, for instance, that the off-diagonal blocks in (A7) are zero.] This implies that the equation for the evolution of W, W^* decouples from that for V, V^*. We have written only Eq. (4.19) for W, W^* since V, V^* are not related to the Stokes parameters.

*Research supported by the National Science Foundation under Grant No. NSF-GP-32996X.

[1]S. Chandrasekhar, *Radiative Transfer* (Dover, New York, 1960).
[2]W. Kohler and G.C. Papanicolaou, Commun. Pure Appl. Math. **27**, 641 (1974).
[3]I.M. Besieris and F.D. Tappert, J. Math. Phys. **13**, 704 (1973).
[4]K.M. Watson, Phys. Fluids **13**, 2514 (1970); J. Math. Phys. **10**, 688 (1968).
[5]C.W. Lau and K.M. Watson, J. Math. Phys. **11**, 3125 (1970).
[6]J. L. Peacher and K.M. Watson, J. Math. Phys. **11**, 3125 (1970).
[7]M.S. Howe, Phil. Trans. Roy. Soc. Lond. **274**, 523 (1973).
[8]R. Bourret, Optica Acta **21**, 721 (1974).
[9]R. Bourret, Astrophys. Space Sci. **24**, 407 (1973).
[10]G.C. Papanicolaou, J. Math. Phys. **13**, 1912 (1972).
[11]W. Kohler and G.C. Papanicolaou, J. Math. Phys. **14**, 1733 (1973).
[12]W. Kohler and G.C. Papanicolaou, J. Math. Phys. **15**, 2186 (1974).
[13]In Ref. 2 it is also shown that, along with weak convergence of $u^\epsilon(\tau)$ to $u^0(\tau)$, moments converge to moments also.
[14]R. Burridge, Proc. Camb. Phil. Soc. **65**, 157 (1969).
[15]N.J. Vilenkin, *Special Functions and the Theory of Group Representations*, Transl. of Math. Monographs, Vol. 22 (AMS, Providence, R.I., 1968).
[16]To avoid confusion with the small parameter ϵ of Sec. 2, the small parameter will be denoted by γ in Secs. 3 and 4.
[17]G.C. Papanicolaou and J.B. Keller, SIAM J. Appl. Math. (Soc. Ind. Appl. Math.) **21**, 287 (1971).
[18]J.B. Keller in *Stochastic Processes in Mathematical Physics anc Engineering*, edited by R. Bellman, Proc. Symp. Appl. Math. **XVI**, 145 (1964).
[19]U. Frisch, in *Probabilistic Methods in Applied Mathematics*, edited by A.T. Bharucha-Reid (Academic, New York, 1968).
[20]G.C. Papanicolaou, and S.R.S. Varadhan, Commun. Pure Appl. Math. **26**, 497 (1973).

Gauge transformations and normal states of the CCR

M. Fannes* and A. Verbeure

Katholieke Universiteit Leuven,† Belgium
(Received 23 May 1975)

Let ω be any normal state on the CCR-algebra, and τ_χ the gauge transformation corresponding to the continuous character χ on the test function space as an Abelian group; denote by K the set of these characters; then we prove that the linear hull of $\{\omega_0\tau_\chi|\chi\in K\}$ is norm dense in all normal states. It is also proved that the theorem is in general false if we take the convex hull.

I. INTRODUCTION

Let (H,σ) be a separable symplectic space, i.e., a real vector space H, equipped with a regular, antisymmetric, real bilinear form σ; and $\overline{\Delta(H,\sigma)}$ the C^*-Weyl algebra built on (H,σ) (see Ref. 1); it is generated by the Weyl functions W:

$$W_\psi(\phi)=0 \quad \text{if} \quad \psi\neq\phi$$
$$=1 \quad \text{if} \quad \psi=\phi$$

with product law $W_\psi W_\phi = W_{\psi+\phi}\exp[-i\sigma(\psi,\phi)]$ and involution $(W_\psi)^*=W_{-\psi}$.

Any operator J on H satisfying $J^*=-J$("+" adjoint with respect to σ), $J^2=-1$ (unit operator), $s_J(\psi,\psi)=-\sigma(J\psi,\psi)>0$ for all $\psi\neq 0$, defines a complex structure on the real space H and a state ω_J on $\overline{\Delta(H,\sigma)}$: $\omega_J(W_\psi)=\exp[-\frac{1}{2}s_J(\psi,\psi)]$. It is a pure quasifree state or Fock state.[2] Consider it GNS-representation π on the Fock space $\mathcal{H}$, and denote by Ω the cyclic vector of the representation of the vacuum state; then $\omega_J(x)=(\Omega,\pi(x)\Omega)$, $x\in\overline{\Delta(H,\sigma)}$. In the following everything will be performed with respect to the representation π; therefore, we drop it and write $\omega_J(x)=(\Omega,x\cdot\Omega)$. Any state ω such that $\omega_\rho(x)=\mathrm{Tr}\rho x$, where ρ is a positive operator such that $\mathrm{Tr}\rho=1$ is called normal state on $\Delta(H,\sigma)$. Furthermore, the real vector space H can be seen as an Abelian group: Let χ be any character of this group; then define τ_χ as a $*$-automorphism of $\overline{\Delta(H,\sigma)}$ by $\tau_\chi(W_\psi)=\chi(\psi)W_\psi$. As τ_χ is isometric with respect to the C^*-norm, it extends to $\overline{\Delta(H,\sigma)}$. These automorphisms are called Gauge transformations. Denote by K the set of all σ-continuous characters, i.e., $\chi\in K$ if there exists a $\psi\in H$ such that $\chi(\phi)=\exp[-2i\sigma(\psi,\phi)]$ for all $\phi\in H$. Then the corresponding Gauge transformation is inner:

$$\tau_\chi(x)=W_\psi x W_{-\psi}, \quad x\in\overline{\Delta(H,\sigma)}.$$

In this paper we prove the following theorem: Let ω be any normal state on the CCR; then any other normal state can in norm be approximated by linear combinations of the states $\omega\cdot\tau_\chi$ with $\chi\in K$.

Also an explicit proof is given of the fact that the theorem is not true if only convex combinations of the gauge transformed state are taken. We have not to stress the usefulness of such a theorem. A particular case is the following: If ω is a quasifree state,[2] then also $\omega\cdot\tau_\chi$ is quasifree. Hence all normal states can in this way be approximated by quasifree states.

In fact the theorem is a generalization of a theorem due to Rocca,[3] also using for the proof a completely different technique. It is a generalization to infinitely many degrees of freedom and the approximative states are not only coherent states, but arbitrary normal states. Our proof yields also a new proof for finite degrees of freedom not using analytic function techniques.

Finally we remark that the result of Lemma II. 2 may be a useful tool in its own, for the study of quasi-equivalence between continuous states on the CCR.

II. THE APPROXIMATION THEOREM

Let $(H_n)_{n\in N}$ be an increasing and absorbing net of finite, J-invariant subspaces of H such that the real dimension of H_n is $2n$. Consider the symplectic basis of H_n: $\{e_i;Je_i|i\in(1,\ldots,n)\}$, i.e., it is a basis of H and $\{e_i|i\in(1,\ldots,n)\}$ is an orthonormal set for s_J and $s_J(e_i,Je_j)=0$ for all i,j. Then any element $\psi\in H$ can be written in the form

$$\psi=\sum_{i=1}^{\infty}(x_ie_i+y_iJe_i), \quad x_i,y_j\in R.$$

By $d\psi$ we denote the Lebesque measure with respect to this basis, i.e., when restricted to H_n

$$d\psi\,|_{H_n}=\prod_{i=1}^{n}dx_i\,dy_i.$$

Furthermore, $\psi_n=\sum_{i=1}^{n}(x_ie_i+y_iJe_i)$ is the projection of ψ on H_n. Finally denote by Ω_ψ the dense set of vectors of $\mathcal{H}$ given by $\Omega_\psi=W_\psi\Omega$; they are also called the coherent state vectors.

The proof of the theorem is performed by a number of steps, formulated in the following lemmas.

Lemma 11.1: With the notations of above, the operators

$$P_n^\psi=(1/\pi)^n\int_{H_n}(\Omega_\psi,W_{-\phi_n}\Omega_\psi)W_{\phi_n}d\phi_n$$

form a decreasing sequence of projections in $\mathcal{B}(\mathcal{H})$ (all bounded operator on $\mathcal{H}$); hence $P^\psi=\lim_n P_n^\psi$ exists, and it is the projection operator on Ω_ψ.

Proof: As the Fock representation is continuous and as

$$\|P_n^\psi\|\leq(1/\pi)^n\int_{H_n}\exp[-s_J(\phi_n,\phi_n)/2]d\phi_n<\infty,$$

by the Lebesque dominated convergence theorem the operators exist in $\mathcal{B}(\mathcal{H})$. Trivially $(P_n^\psi)^*=P_n^\psi$. Now we prove that for $m\geq n$ then $P_n^\psi P_m^\psi=P_m^\psi$. Let us calculate

$$\pi^{n+m}P_n^\psi P_m^\psi=\int_{H_m}d\xi_m(\Omega_\psi;W_{-\xi_m}\Omega_\psi)I(\xi_m)W\xi_m,$$

where

$$I(\xi_m)=\int_{H_n}d\phi_n(\Omega_\psi,W_{-\phi_n}\Omega_\psi)\exp[i\sigma(\xi_m,\phi_n)]$$

$$= \int_{H_n} d\phi_n \exp[-(\phi_n, \phi_n) + i\sigma(\xi_m, \phi_n) - s_J(\xi_m, \phi_n)]$$

$$= \pi^n.$$

This proves that P_n^ψ is a decreasing sequence of projections. Furthermore, analogous calculations yield

$$P_n^\psi \Omega_\psi = \Omega_\psi \quad \text{for all } n \in N,$$

$$P^\psi W_\phi \Omega_\phi = (\Omega_\psi, W_\phi \Omega_\phi)\Omega_\psi \quad \text{for all } \phi \in H. \qquad \text{QED}$$

Lemma II. 2: Let A be a bounded operator on $\mathcal{H}$, i.e., $A \in \mathcal{B}(\mathcal{H})$, and ω any normal state on $\overline{\Delta(H, \sigma)}$; then the function on H

$$\psi \to A(\psi) = [1/\omega(W_\psi)] \int d\phi\, \omega(W_{-\phi} A W_\phi) \exp[2i\sigma(\psi, \phi)]$$

is square integrable if and only if A is of Hilbert–Schmidt class in $\mathcal{B}(\mathcal{H})$ and

$$\mathrm{Tr} A^* A = \int d\tilde{\psi} |A(\psi)|^2, \qquad (*)$$

where $d\tilde{\psi}$ stands for $(1/\pi)^n d\psi_n$ when restricted to H_n.

Proof: As the state ω is normal, the map $\phi \in H \to \omega(W_{-\phi} A W_\phi) \cdot \exp[2i\sigma(\psi, \phi)]$ is measurable and the function $A(\psi)$ is defined for all $A \in \mathcal{B}(\mathcal{H})$. As the finite rank operators are dense in the Hilbert–Schmidt operators and as the set of vectors $\{\Omega_\psi, \psi \in H\}$ is a total set in $\mathcal{H}$, it is sufficient to prove the relation $(*)$ for $A = P^\psi$, where P^ψ are the projection operators of Lemma II. 1. A straightforward calculation shows

$$A(\psi) = [1/\omega(W_\psi)] \int_H d\phi\, \omega(W_{-\phi} A W_\phi) \exp[2i\sigma(\psi, \phi)]$$

$$= \mathrm{Tr} A W_{-\psi}$$

and

$$\mathrm{Tr} A^* A = \int_H d\tilde{\xi} |\mathrm{Tr} A W_{-\xi}|^2;$$

hence $\mathrm{Tr} A^* A = \int_H d\tilde{\psi} |A(\psi)|^2$ QED

Remark: The lemma remains true if we replace the condition "ω a normal state" by "ω a continuous state."

Lemma II. 3: Let ρ be any density matrix in $\mathcal{B}(\mathcal{H})$; then the set of trace class operators $\{W_\psi \rho W_{-\psi} | \psi \in H\}$ is trace norm dense in the set of all trace class operators.

Proof: The set θ_1 of trace class operators is an ideal of $\mathcal{B}(\mathcal{H})$, and θ_1 equipped with the trace norm a Banach space; the topological dual θ_1^* is $\mathcal{B}(\mathcal{H})$, and we prove that the topological dual of the set $\{W_\psi \rho W_{-\psi} | \psi \in H\}$ is also $\mathcal{B}(\mathcal{H})$, by showing that this set of operators is separating for $\mathcal{B}(\mathcal{H})$. Therefore, let $S \in \mathcal{B}(\mathcal{H})$ such that

$$\mathrm{Tr} S W_\phi \rho W_{-\phi} = 0 \quad \text{for all } \phi \in H$$

or, in terms of the state $\omega_\rho(x) = \mathrm{Tr}\rho x$, this is equivalent to $\omega_\rho(W_{-\phi} S W_\phi) = 0$ for all $\phi \in H$, yielding, in the notation of Lemma II. 2, $S(\psi) = 0$; hence $\mathrm{Tr} S^* S = 0$ and $S = 0$.

QED

Theorem II. 4: Let ω_ρ be any normal state on $\overline{\Delta(H, \sigma)}$; then the linear hull of the set of linear functionals generated by $\{\omega_\rho \cdot \tau_\chi | \chi \in K\}$ is norm dense in the set of normal state on $\overline{\Delta(H, \sigma)}$.

Proof: As ω_ρ is a normal state, it can be written as

$$\omega_\rho(x) = \mathrm{Tr}\rho x : x \in \mathcal{B}(\mathcal{H}),$$

where ρ is a density matrix (i.e., $\rho \geq 0$, $\mathrm{Tr}\rho = 1$) and

$$\omega_\rho(\tau_\chi(x)) = \mathrm{Tr}\rho\tau_\chi(x) = \mathrm{Tr}\rho W_\psi x W_{-\psi} = \mathrm{Tr} W_{-\psi} \rho W_\psi x.$$

Also as $\mathrm{Tr} W_{-\psi} \rho W_\psi = \mathrm{Tr}\rho = 1$, the gauge transformations map normal states into normal states or, stated otherwise, map the set of density matrices into itself. Furthermore, the well-known relation

$$\|\omega_{\rho_1} - \omega_{\rho_2}\| = \sup_{x \in \mathcal{B}(\mathcal{H})} [|\mathrm{Tr}(\rho_1 - \rho_2)(x)| / \|x\|]$$

$$= \mathrm{Tr}[(\rho_1 - \rho_2)^2]^{1/2} = \mathrm{Tr}|\rho_1 - \rho_2|$$

connects the norm topology on the normal states with the trace norm on the density matrices. From this it follows that the statement of the theorem is equivalent to the statement of Lemma II. 3 and hence the theorem follows. QED

In the following theorem we show that it is not true that the convex hull of $\{\omega_\rho \cdot \tau_\chi | \chi \in K\}$, where ω_ρ is any normal state, is norm dense in the set of normal states. The proof is only given for $\omega_\rho = \omega_J$ (the Fock state). A special case of this situation is known in the domain of quantum theory of optical coherence in the following form. Let H_n be finite dimensional then in the so-called P-representation[10] of the density matrix ρ:

$$\rho = \int_{H_n} P^\psi P(\psi)\, d\psi,$$

where P^ψ is defined in Lemma II. 1; it is known that the function $\psi \to P(\psi)$ is not necessarily positive definite.[5] Explicit examples of the function $P(\psi)$ are known with $P(\psi)$ negative on sets of nonzero measure.[6]

Theorem II. 5: Let ω_J be the Fock state on $\overline{\Delta(H, \sigma)}$; then there exists a normal state ω_0 which do not belong to norm closure of the convex hull of $\{\omega_J \cdot \tau_\chi | \chi \in K\}$

Proof: Let A_i be the element of $\mathcal{B}(\mathcal{H})$, defined by

$$A_i = \tfrac{1}{2} \int_{-\infty}^{\infty} dy \exp(-|y|) W y J e_i$$

for a fixed index i. By spectral theory one gets

$$A_i = 1/(1 + P_i^2),$$

where P_i is the momentum operator (i-component) defined by

$$P_i = \text{s-}\lim_{y \to 0} [(W y J e_i - 1)/y].$$

Hence, spectrum $A_i = [0, 1]$. Furthermore,

$$\omega_J(\tau_\chi(A_i)) = \tfrac{1}{2} \int_{-\infty}^{\infty} dy \exp(-|y|) \chi(y J e_i) \exp(-y^2/4)$$

and

$$\sup_\chi \omega_J(\tau_\chi(A_i)) = \gamma < 1;$$

therefore, for all ω in the convex hull of $\{\omega_J \cdot \tau_\chi | \chi \in K\}$

$$\omega(A_i) \leq \gamma.$$

However, as spectrum $A_i = [0, 1]$ there exists a normal state ω_0 such that

$$\gamma < \omega(A_i) \leq 1;$$

therefore, ω_0 does not belong to the norm closure of the convex hull of $\{\omega_J \cdot \tau_\chi | \chi \in K\}$. QED

III. DISCUSSION

About the techniques used let us remark the following.

The projection operator of Lemma II. 1 is a special case of a more general result which is easily proved at least for finite dimensions, namely let ω be a normal state; then $\omega(x) = \mathrm{Tr}\rho x$, $x \in \beta(H)$ and the density matrix is given by

$$\rho = (1/\pi)^n \int_H d\psi\ \omega(W_{-\psi})W_\psi,$$

where the integral is in the strong operator topology.

This formula answers, by an explicit construction, the question: given a state ω, what is its density matrix? Note the difference with the map $\psi \to A(\psi)$ of Lemma II. 2, which is essentially the Wigner phase-space quasi-probability distribution, as studied by Moyal,[7] Pool,[8] Loupias and Miracle-Sole[9] for finitely many degrees of freedom ($H = R^n + R^n$). In Lemma II. 2 this condition is removed, and this should encourage the possibility of studying the Weyl correspondence and Wigner quasi-probability density function in quantum field theory. Also the map of Lemma II. 2 is related to the "reproducing kernel" of McKenna and Klauder[4] in their study of representations of the canonical commutation relation.

Finally let us point out the physical relevance of the theorem. As said before in the Introduction, the theorem is a generalization of the optical equivalence theorem, widely used in quantum optics. We do not insist more on this point, but refer to the litterature; see, e.g., Refs. 3.

It is a generalization in the sense that the approximative states are not only coherent states, but arbitrary normal states. In quantum mechanics we think about the treatment of perturbed solvable models. The global solution may be written in terms of the solutions of the solvable model, which has not to be a coherent state. As it is a generalization to infinitely many degrees of freedom, its usefulness in statistical mechanics is manifestly clear.

ACKNOWLEDGMENT

We thank F. Rocca for a fruitful discussion.

*Aspirant NFWO, Belgium.
†Postal Address: Instituut voor Theoretische Fysica, Celestijnenlaan 200 D, B-3030, Heverlee, Belgium.

[1] J. Manuceau, M. Sirugue, D. Testard, and A. Verbeure, Comm. Math. Phys. **32**, 231 (1973).
[2] J. Manuceau and A. Verbeure, Comm. Math. Phys. **9**, 293 (1968).
[3] F. Rocca, C. R. Acad. Sci. Paris A **262**, 547 (1966); J.R. Klauder and E.C.G. Sudarshan, *Quantum Optics* (Benjamin, New York, 1968).
[4] J.R. Klauder and J. McKenna; J. Math. Phys. **6**, 68 (1965) and references cited therein.
[5] R.J. Glauber *Fundamental Problems in Statistical Mechanics. II*, edited by E.G.D. Cohen (North-Holland, Amsterdam, 1968), p. 140.
[6] F. Rocca, private communication.
[7] J.E. Moyal, Proc. Cambridge Phil. Soc. **45**, 99 (1949).
[8] J.C.T. Pool, J. Math. Phys. **7**, 66 (1966).
[9] Y. Loupias and S. Miracle-Sole, Comm. Math. Phys. **2**, 31 (1966).
[10] R.J. Glauber, Phys. Rev. **131**, 2766 (1963).

New class of exact solutions of the Dirac equation

Franklin S. Felber and John H. Marburger

Department of Physics, University of Southern California, Los Angeles, California 90007
(Received 3 February 1975)

A new class of exact solutions of the Dirac equation with external electromagnetic fields is derived by assuming a set of field-dependent solution matrices which obey an algebra isomorphic to the Pauli matrices. The method of exact solution may be applied to any field having a four-vector potential A^μ depending only on $k^\mu x_\mu$, but for which the field tensor and initial electron momentum are such that $A^\mu A_\mu$, $A^\mu p_\mu$, $(\sigma^{\mu\nu}F_{\mu\nu})^2$, and $(\sigma^{\mu\nu}F'_{\mu\nu})^2$ are independent of $k^\mu x_\mu$. Exact solutions for a circularly polarized propagating electromagnetic wave in an isotropic medium, for a screw symmetric static magnetic field, and for a rotating uniform electric field are given in terms of the roots of a quartic equation. A class of solutions is given explicitly in the weak field limit to lowest order in eA/mc^2. The vacuum limit of the solution of a wave propagating in a medium is shown to be the Volkov solution.

I. INTRODUCTION

Exact solutions of the Dirac equation have been found for only a few external field configurations. The most important exact solutions known are those for the Coulomb potential,[1] homogeneous electric and magnetic fields,[2,3] and the field of a plane wave in a vacuum.[4] In this paper we derive a new class of exact solutions which is applicable to a variety of fields, including electromagnetic waves for which the length of the wave four-vector is nonzero.

In the next section we outline the conditions which must be satisfied by fields which allow exact solutions by the techniques of this paper. The exact class of solutions is derived in Sec. III, and written in terms of the roots of a quartic equation having two field dependent parameters. The parameters are given explicitly for a circularly polarized electromagnetic field in a medium with refractive index $n>1$, for a magnetic field with screw symmetry, and for a rotating electric field.

In Sec. IV we derive explicitly several limiting cases of the exact solutions given in Sec. III. In the vacuum limit we treat a circularly polarized propagating wave in a medium to retrieve the Volkov solutions. We also present the "weak field" limit of the solutions valid for $eA/mc^2 \ll 1$.

II. THE CLASS OF SOLUTIONS DEFINED

In the following we shall use the metric $g^{00}=-g^{ii}=1$, and the scalar product of the four-vectors x^μ and y_ν will be denoted by xy. We shall set $\hbar=c=1$. The four-potential A^μ is taken to depend on the four-coordinates only in the combination $\phi=kx$, where k^μ is the "wave four-vector." A^μ shall satisfy the Lorentz gauge condition $\partial A=A'k=0$, and the field tensor can be written $F_{\mu\nu}=A'_\mu k_\nu - A'_\nu k_\mu$. A prime denotes differentiation with respect to the "phase ϕ."

All matrices are 4×4. We define an antisymmetric tensor $\sigma^{\mu\nu}$ having components

$$\sigma^{ij}=\begin{bmatrix}\sigma^k & 0\\ 0 & \sigma^k\end{bmatrix}$$

with $i,j,k=1,2,3$ in cyclic order, and

$$\sigma^{0k}=i\alpha^k=i\begin{bmatrix}0 & \sigma^k\\ \sigma^k & 0\end{bmatrix}.$$

The σ^k are the usual 2×2 Pauli matrices.

In order for the Dirac equation to be exactly solvable by the methods of the following section, the four-potential of the field must satisfy the following conditions:

A^2, Ap, $(\sigma^{\mu\nu}F_{\mu\nu})^2$, and $(\sigma^{\mu\nu}F'_{\mu\nu})^2$ must all be independent of ϕ.

Throughout this paper p^μ denotes a constant four-vector satisfying the condition $p^2=m^2$. When the field is removed, p^μ becomes the free particle four-momentum.

In all that follows we shall take the following as our particular choice of particle motion and field:

$$\begin{aligned}&A^\mu(\phi)=a(0,\cos\phi,\sin\phi,0),\quad a=\text{const.},\\ &p^\mu=(p_0,0,0,p_3),\quad k^\mu=(k_0,0,0,k_3).\end{aligned}\tag{1}$$

This choice satisfies the conditions above, and allows for a variety of interesting solutions, but does not exhaust the number of fields for which exact solutions may be found by the methods of this paper.

III. A FORMAL DERIVATION OF THE CLASS OF SOLUTIONS

The Klein–Gordon equation,

$$(-\partial^2-2ieA\partial+e^2A^2-m^2)Z_0=0\tag{2}$$

is known to have the solution[5]

$$Z_0=\exp(iS),\tag{3}$$

where the classical action

$$S=-px+(kp-R)\phi/k^2\tag{4}$$

is the solution of the Hamiltonian–Jacobi equation.

The useful constants R and b are defined as

$$\begin{aligned}R&=kp\{1-[k^2/(kp)^2][(p-eA)^2-m^2]\}^{1/2}\\ &=kp[1+e^2a^2k^2/(kp)^2]^{1/2},\\ b&=-iR(k^2)^{-1},\end{aligned}\tag{5}$$

and the positive root is taken.

The Volkov solutions to the Dirac equation are of the form $Y_0 \exp(iS_0)$, where Y_0 is a field dependent matrix front factor, and S_0 is the classical action for an electromagnetic field in a vacuum. We therefore assume the form $Y(\phi) Z_0$ for our solutions, with Z_0 defined in (3). An equation satisfied by $Y(\phi)$ can be obtained from the second order Dirac equation[6]

$$(-\partial^2 - 2ieA\partial + e^2A^2 - m^2 + \tfrac{1}{2}e\sigma^{\mu\nu}F_{\mu\nu})\, YZ_0 = 0. \tag{6}$$

Direct substitution of (3) and (4) leads to

$$Y''(\phi) + 2bY'(\phi) - f\Omega(\phi)\, Y(\phi) = 0, \tag{7}$$

where f is a constant scalar

$$f \equiv iea(k^2)^{-1/2}, \tag{8}$$

and Ω is the 4×4 matrix

$$\Omega(\phi) \equiv (fk^2)^{-1}\, \tfrac{1}{2}e\sigma^{\mu\nu}F_{\mu\nu} \tag{9}$$

or

$$\Omega(\phi) = (k^2)^{-1/2}(\alpha_2 k_0 - i\sigma^{23}k_3)\exp(i\sigma^{12}\phi) \tag{9'}$$

for our choice of field (1).

Equation (7) can be solved by assuming that Y is a linear combination of the four matrices

$$\mathbb{1}, \sigma^{12}, \Omega(\phi), \Omega'(\phi),$$

whose algebraic properties are isomorphic to those of the 2×2 Pauli matrices: They form a closed set under multiplication (to within a constant phase factor), the square of each is $\mathbb{1}$, and all but the unit matrix are traceless and anticommute with each other. The matrix Ω' is

$$\Omega'(\phi) = \partial\Omega/\partial\phi = i\Omega\sigma^{12}. \tag{10}$$

Letting

$$Y(\phi) = t(\phi)\,\mathbb{1} + u(\phi)\,\sigma^{12} + v(\phi)\,\Omega(\phi) + w(\phi)\,\Omega'(\phi) \tag{11}$$

and using (10), the anticommutation relations

$$\{\sigma^{12}, \Omega\} = \{\sigma^{12}, \Omega'\} = \{\Omega, \Omega'\} = 0, \tag{12}$$

and the relation

$$\Omega'' = i\Omega'\sigma^{12} = -\Omega, \tag{13}$$

we substitute (11) into (7) to obtain four coupled *scalar* equations in the four unknown scalar functions of ϕ:

$$\begin{aligned} &t'' + 2bt' - fv = 0,\\ &u'' + 2bu' - ifw = 0,\\ &v'' + 2bv' - v - 2w' - 2bw - ft = 0,\\ &w'' + 2bw' - w + 2v' + 2bv + ifu = 0. \end{aligned} \tag{14}$$

The solutions of the first order Dirac equation are easily obtained from the solutions of the second order equation. Therefore, finding a class of exact solutions to the Dirac equation is contingent only on the ability to solve the system of equations (14) exactly.

We proceed to solve (14) exactly by defining new variables

$$\eta = t + u, \quad \xi = t - u, \quad \nu = v + iw, \quad \mu = v - iw. \tag{15}$$

Equations (14), written in terms of the new variables, become

$$\begin{aligned} &\eta'' + 2b\eta' = f\nu, \quad \xi'' + 2b\xi' = f\mu,\\ &(\nu\exp i\phi)'' + 2b(\nu\exp i\phi)' = f(\eta\exp i\phi),\\ &[\mu\exp(-i\phi)]'' + 2b[\mu\exp(-i\phi)]' = f[\xi\exp(-i\phi)]. \end{aligned} \tag{16}$$

These equations are equivalent to two uncoupled, homogeneous, linear fourth order equations in η and ξ, or alternatively in ν and μ. The solutions are

$$\begin{aligned} &\eta = \sum_{j=1}^{4} C_j \exp(r_j\phi), \quad \xi = \sum_{I=1}^{4} D_I \exp(r_I^*\phi),\\ &\nu = f^{-1}\sum_{j=1}^{4} C_j(r_j^2 + 2br_j)\exp(r_j\phi),\\ &\mu = f^{-1}\sum_{I=1}^{4} D_I(r_I^{*2} + 2br_I^*)\exp(r_I^*\phi), \end{aligned} \tag{17}$$

where the r_j are the roots of the quartic equation

$$\begin{aligned} &r^4 + 2(2b+i)r^3 + (4b^2 + 6ib - 1)r^2\\ &\quad + 2b(2bi - 1)r - f^2 = 0 \end{aligned} \tag{18}$$

and the r_I^* are the roots of the same equation with the explicit i replaced everywhere by $-i$. (This is not the complex conjugate equation, since b and f can be imaginary.)

There are eight arbitrary constants, C_j and D_I, corresponding to the eight independent solutions of the second order equation. The solutions $\psi(\phi)$ of the first order Dirac equation,

$$[\gamma(i\partial - eA) - m]\,\psi(\phi) = 0, \tag{19}$$

are given in terms of the solutions $Z(\phi)$ of the second order equation by

$$\psi(\phi) = [\gamma(i\partial - eA) + m]\, Z(\phi). \tag{20}$$

Thus the most general exact solutions of the first order equation of the class considered are

$$\begin{aligned} \psi(\phi) = [\gamma(i\partial - eA) + m]\Big[&\left(\frac{\eta+\xi}{2}\right)\mathbb{1} + \left(\frac{\eta-\xi}{2}\right)\sigma^{12}\\ &+ \left(\frac{\nu+\mu}{2}\right)\Omega + \left(\frac{\nu-\mu}{2i}\right)\Omega'\Big]\exp(iS). \end{aligned} \tag{21}$$

It is convenient to choose the arbitrary constants C_i and D_I so that when $\phi = 0$, the solution resembles a free particle solution of the first order Dirac equation. This choice implies the eight conditions

$$Z(0) = \mathbb{1}\exp(-ipx) \tag{22a}$$

and

$$\psi(0) = [\gamma(p - eA) + m]\, Z(0). \tag{22b}$$

These "initial" conditions impose in turn eight conditions on the constants C_j and D_I (all summations are from 1 to 4):

$$\sum C_j = 1, \qquad \sum D_I = 1, \tag{23a}$$

$$\sum C_j r_j = 0, \qquad \sum D_I r_I^* = 0, \tag{23b}$$

$$f^{-1}\sum C_j r_j(r_j + 2b) = 0, \qquad f^{-1}\sum D_I r_I^*(r_I^* + 2b) = 0 \tag{23c}$$

$$(\gamma k\Omega) f^{-1} \sum C_j r_j (r_j + 2b)(r_j + i) = 0,$$

$$(\gamma k\Omega) f^{-1} \sum D_l r_l^* (r_l^* + 2b)(r_l^* - i) = 0. \tag{23d}$$

Now for any field which satisfies (1), we need only specify the values of the parameters b and f in (18) in order to have the exact solutions for a Dirac electron in that field.

Consider for example $k_3 = nk_0$. Then for the field and initial electron motion given by (1), the solutions (21), together with conditions (17) and (23), represent the exact solution for a Dirac electron emitted in the direction of propagation of a right circularly polarized electromagnetic wave in a transparent, isotropic medium of index of refraction n. For this case the values of the parameters b and f are

$$b_n = \frac{i(p_0 - np_3)}{k_0(n^2-1)} \left(1 - \frac{e^2a^2(n^2-1)}{(p_0 - np_3)^2}\right) \tag{24}$$
$$f_n = eak_0^{-1}(n^2-1)^{-1/2}.$$

Or we may consider $k_0 = 0$ in (1). Then (17), (21), and (23) represent the exact solution for a Dirac electron emitted parallel to the screw symmetry axis of the static magnetic field $\mathbf{B} = ak_3(\hat{\imath}\cos k_3x_3 - \hat{\jmath}\sin k_3x_3)$.

For this case b and f take on values

$$b_B = -ip_3k_3^{-1}[1 - (ea/p_3)^2]^{1/2} \text{ and } f_B = eak_3^{-1}. \tag{25}$$

Similarly if we let $k_3 = 0$ in (1), then (17), (21), and (23) represent the solution for a Dirac electron emitted parallel to the rotation axis of a rotating uniform electric field. The relevant parameters for this situation are

$$b_{\mathcal{E}} = -ik_0^{-1}[p_0^2 + (ea)^2]^{1/2} \text{ and } f_{\mathcal{E}} = ieak_0^{-1}. \tag{26}$$

IV. LIMITING CASES OF THE EXACT CLASS OF SOLUTIONS

The solution of the quartic equation (18) is not trivial in general[7] and, once found, does not provide much in the way of new physical insights. There are, however, several interesting limits of the exact solutions. We consider in this section the weak field limit for the entire class of solutions and the vacuum limit for a propagating electromagnetic field in a medium.

A. Weak field limit

In dimensional units, the parameter f has magnitude $|f| = ea/\hbar ck_0\theta$, where $\theta = (k^2)^{1/2}/k_0$. This may be written as

$$\theta|f| = E/E_d, \tag{27}$$

in which E is the peak amplitude of the electric field associated with a and E_d is a scaling field. To be specific, we shall assume $k_3 = nk_0$. The energy density corresponding to E_d in cgs Gaussian units is

$$(n^2+1)E_d^2/8\pi = (\hbar c/e^2)\cdot 2\pi^2 \cdot \hbar\omega\lambda_0^{-3}(n^2+1)$$
$$= 5.375\times10^{-13}(n^2+1)/\lambda_0^4, \tag{28}$$

where $\lambda_0 \equiv 2\pi/k_0 = 2\pi c/\omega$. The corresponding power density for $\lambda_0 = 1$ μm is 8×10^6 watt/cm^2, which is a modest field strength with current laser technology. We find below, however, that to lowest order in f only the combination f/b appears, which has magnitude (for $|p_3| \ll mc$)

$$|f/b\theta| \approx (eE/m\omega c). \tag{29}$$

This dimensionless quantity is ubiquitous in treatments of the motion of charged particles in intense plane radiation fields.[8] When $eE/m\omega c$ approaches or exceeds unity, the particle motion is relativistic. Thus the weak field limit below is equivalent to the nonrelativistic limit. The power density at $\lambda_0 = 1$ μm corresponding to the field $E_R = m\omega c/e$ is 8.7×10^{17} watt/cm^2.

In the limit $|\epsilon| \ll 1$, where $\epsilon = f/2b$, we shall work to order ϵ^2, but for simplicity present the solutions only to order ϵ.

To order ϵ^2 then, the constant b and the action take on the values

$$b \approx (-ikp/k^2)(1 + 2\epsilon^2), \quad S \approx -px - 2ib\epsilon^2\phi. \tag{30}$$

The roots of the quartic equation (18) and its sister equation to the same order are

$$\begin{aligned} r_1 &= \Delta_1, \quad r_2 = -2b + \Delta_2, \\ r_3 &= -i - \Delta_2, \quad r_4 = -2b - i - \Delta_1, \\ r_1^* &= -\Delta_2, \quad r_2^* = -2b - \Delta_1, \\ r_3^* &= i + \Delta_1, \quad r_4^* = -2b + i + \Delta_2, \end{aligned} \tag{31}$$

where

$$\Delta_1 = 2b\epsilon^2/(2ib - 1) \text{ and } \Delta_2 = 2b\epsilon^2/(2ib + 1). \tag{32}$$

With these values for the roots, the initial conditions (23) require the constants to be

$$\begin{aligned} C_1 &= 1 - \left(\frac{8b^2(1-2ib) + 1 - 4ib}{(1+4b^2)2b}\right)\Delta_1, \\ C_2 &= \left(\frac{1-2ib}{1+2ib}\right)\frac{\Delta_1}{2b}, \\ C_3 &= \left(\frac{2b+i}{1+2ib}\right)\Delta_1, \\ C_4 &= \left(\frac{-i}{1-2ib}\right)\Delta_1, \\ D_1 &= 1 + \left(\frac{8b^2(1+2ib) + 1 + 4ib}{(1+4b^2)2b}\right)\Delta_2, \\ D_2 &= -\left(\frac{1+2ib}{1-2ib}\right)\frac{\Delta_2}{2b}, \\ D_3 &= -\left(\frac{2b-i}{1-2ib}\right)\Delta_2, \\ D_4 &= \left(\frac{-i}{1+2ib}\right)\Delta_2. \end{aligned} \tag{33}$$

Thus the solution of the Dirac equation to order ϵ, for any field satisfying (1) and (23), is

$$\psi(\phi) \approx [\gamma(p - eA) + m](\mathbb{1} + \Delta)\exp(iS), \tag{34a}$$

where

$$\begin{aligned} \Delta = -\tfrac{1}{2}b\,\{&(2ib-1)^{-1}[2ib + e^{-i\phi}(1 - 2ib - e^{-2b\phi})] \\ &\times(\Omega' + i\Omega) + (2ib+1)^{-1}[2ib - e^{i\phi}(1 + 2ib - e^{-2b\phi})] \\ &\times(\Omega' - i\Omega)\}. \end{aligned} \tag{34b}$$

B. Vacuum limit

In the vacuum limit we shall consider $k^\mu = k_0(1, 0, 0, n)$, and let $k^2(ea)^{-2}$ approach zero by virtue of $n \to 1$, and not by $k_0 \to 0$. Then f^2 and b both approach infinity as $(k^2)^{-1}$, and the action approaches the limiting value

$$S \to -px - (e^2a^2/2kp)\,\phi. \tag{35}$$

Keeping only terms of the highest order in $(k^2)^{-1}$, we find the roots of (18) and its sister equation to be

$$\begin{aligned} &r_1 = -i\epsilon^2, \quad r_2 = -i, \quad r_3 = -2b, \\ &r_1^* = i\epsilon^2, \quad r_2^* = i, \quad r_3^* = -2b. \end{aligned} \tag{36}$$

There are only six unique roots corresponding to six independent solutions. Fortunately, there are only six initial conditions as well. Since $\gamma k \Omega$ vanishes in the vacuum limit, Eqs. (23d) do not represent conditions on the constants in this limit.

Substitution of (36) into (23) yields

$$C_1 = D_1 = 1 + \epsilon^2, \quad C_2 = D_2 = -\epsilon^2, \quad C_3 = D_3 = 0. \tag{37}$$

Thus the exact solution of the Dirac equation for an electron emitted parallel to the direction of propagation of a right circularly polarized electromagnetic wave in a *vacuum* is found to agree with the Volkov solution[4]:

$$\begin{aligned} \psi_{\text{vac}} &= [\gamma(p - eA) + m]\{\mathbb{1} - \epsilon[\Omega'(1 - \cos\phi) - \Omega\sin\phi]\}\exp(iS) \\ &= [\gamma(p - eA) + m]\{\mathbb{1} + [ea(k^2)^{1/2}/2kp][\Omega'(\phi) - \Omega'(0)]\} \\ &\quad \times \exp(iS), \end{aligned} \tag{38}$$

where S is given by (35).

This solution could have been obtained directly from (34) as well by simply letting b approach infinity in the expression for Δ.

ACKNOWLEDGMENT

The authors are grateful to Professor J. S. Nodvik for useful discussions.

[1] C. G. Darwin, Proc. Roy. Soc. (London) A 118, 654 (1928); W. Gordon, Z. Phys. 48, 11 (1928).
[2] I. I. Rabi, Z. Phys. 49, 7 (1928).
[3] References to other exact solutions may be found in Lui Lam, Can. J. Phys. 48, 1935 (1970).
[4] D. M. Volkov, Z. Phys. 94, 250 (1935).
[5] A. S. Dement'ev, A. G. Kul'kin, and Yu. G. Pavlenko, Zh. Eksp. Teor. Fiz. 62, 161 (1972) [Sov. Phys., JETP 35, 86 (1972)].
[6] V. B. Berestetskii, E. M. Lifshitz, and L. P. Pitaevskii, *Relativistic Quantum Theory, Part I*, translated by J. B. Sykes and J. S. Bell (Addison-Wesley, Reading, Mass., 1971).
[7] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions*, National Bur. Standards Appl. Math. Series 55, (U. S. Gov. Printing Office, Washington, D. C., 1964), p. 17.
[8] J. H. Eberly, in *Progress in Optics*, edited by E. Wolf (North-Holland, Amsterdam, 1969), Vol. VII.

On the determination of the relativistic wave equations associated with a given representation of $SL(2,C)$

William J. Hurley and E. C. G. Sudarshan

Center for Particle Theory, University of Texas, Austin, Texas 78712
(Received 21 March 1974; revised manuscript received 2 August 1974)

Straightforward algebraic techniques are presented and used to determine the structure of wave equations whose relativistic covariance is governed by two representations of $SL(2,C)$, $S_0(\Lambda) = (1,1/2) \oplus (1/2,1) \oplus (1/2,0) \oplus (0,1/2)$ and $S_1(\Lambda) = S_0(\Lambda) \oplus (1/2,0) \oplus (0,1/2)$, subject to the requirements that the equations should be parity preserving, admit an invariant Hermitian bilinear form realized by a numerical matrix η, and that they should describe a particle with a unique mass and spin. It is shown that $S_0(\Lambda)$ leads to a unique algebraic structure, that of the Rarita–Schwinger equation, whereas $S_1(\Lambda)$ leads either to a trivial extension of the former case or to a family of equations whose matrices have a minimal algebra with degree one higher than that of the former case. One such example reproduces the equation presented by Glass. When, contrary to custom, a singular η matrix is considered, it is shown that $S_1(\Lambda)$ allows for equations whose coefficient matrices are reducible but indecomposable. These equations are completely equivalent to the Rarita–Schwinger equation in the free case, but the added components may enter the dynamics in the presence of certain interactions. The present examples serve to illustrate techniques which may be applied in the study of any relativistic wave equation.

I. INTRODUCTION

A relativistic wave equation describing free, massive ($m > 0$) particles may be written in the general form

$$(i\beta \cdot \partial - m)\,\phi(x) = 0 \tag{I. 1}$$

where β_μ represents four $N \times N$ numerical matrices with the property that

$$S(\Lambda)\beta_\mu\, S(\Lambda^{-1}) = (\Lambda\beta)_\mu \tag{I. 2}$$

where $S(\Lambda)$ is the N-dimensional representation of $SL(2, C)$ which governs the index transformation of ϕ under the Poincaré group **P**,

$$\phi'_\alpha(x) = S_{\alpha\beta}(\Lambda)\,\phi_\beta(\Lambda^{-1}x + a), \quad \alpha, \beta = 1, \ldots, N, \tag{I. 3}$$

Λ being the homogeneous Lorentz transformation and a the space–time translation.

In a recent study[1] straightforward constructive algebraic techniques[2] were used in order to classify under very general conditions the possible $S(\Lambda)$ which may be associated with β_μ which satisfied a given algebraic condition. A large class of such $S(\Lambda)$ was thereby enumerated. In the present note we shall employ these same constructive techniques in order to determine all possible β_μ associated with a given $S(\Lambda)$. More specifically, we shall examine the representations

$$S_0(\Lambda) \equiv (1, \tfrac{1}{2}) \oplus (0, \tfrac{1}{2}) \oplus (\tfrac{1}{2}, 1) \oplus (\tfrac{1}{2}, 0) \tag{I. 4}$$

and

$$S_1(\Lambda) \equiv S_0(\Lambda) \oplus (\tfrac{1}{2}, 0) \oplus (0, \tfrac{1}{2}) \tag{I. 5}$$

and seek all possible β_μ which are covariant, parity symmetric, yield a unique spin-3/2, a unique mass $m > 0$, and permit the existence of a Hermitianizing matrix η.

We shall find that these general restrictions determine a unique β-algebra in the case of $S_0(\Lambda)$ and a very limited number of possibilities for $S_1(\Lambda)$. The Rarita–Schwinger (R–S) equation[3] and the Glass equation[4,5] are recovered in addition to other related possibilities. By relaxing the requirement that η be nonsingular we are led to a new class of equations for $S_1(\Lambda)$ which have reducible but indecomposable β matrices. These equations are completely equivalent to the R–S theory in the free case but may deviate from the latter's predictions in the presence of interactions. We refer to the additional components as "barnacles."

In the next section we shall present our basic assumptions and apply them to $S_0(\Lambda)$ and $S_1(\Lambda)$ in Secs. III and IV, respectively, giving a complete listing of the possible wave equations associated with each representation. In Sec. V we shall briefly discuss the physical significance of the various possible wave equations and of the new barnacled equations.

II. ASSUMPTIONS

We shall impose the following restrictions upon the wave equation (I. 1):

(1) Equation (I. 1) is covariant under the proper Poincaré group and hence the $S(\Lambda)$ and β_μ must satisfy Eq. (I. 2). In terms of the generators of rotations and boosts, J_i and N_i, in the representation $S(\Lambda)$ this requirement may be stated entirely in terms of β_0 as

$$[J_i, \beta_0] = 0, \tag{II. 1a}$$

$$[[N_3, \beta_0], N_3] = \beta_0. \tag{II. 1b}$$

(2) Equation (I. 1) is covariant under space reflection, i. e., there exists a numerical matrix P such that $[x' = (x_0, -\mathbf{x})]$

$$\phi'(x') = \alpha_p P \phi(x) \tag{II. 2}$$

where α_p is a phase, $P^2 = I$, and

$$[\beta_0, P]_- = 0, \tag{II. 3a}$$

$$[\beta_i, P]_+ = 0, \quad i = 1, 2, 3. \tag{II. 3b}$$

Such a matrix will exist only if $S(\Lambda)$ is self-conjugate, i. e., composed of representations of the form (n, n) or $(n, m) \oplus (m, n)$. It will have matrix elements which are

nonvanishing only in those blocks which connect conjugate representations and there it will be a multiple of the identity. For the two representations $S_0(\Lambda)$ and $S_1(\Lambda)$ we fix the relative scale of the conjugate components such that

$$P=\begin{array}{c|cc} & A_i & A_i^c \\ \hline A_i & 0 & 1 \\ A_i^c & 1 & 0 \end{array}, \qquad i=0,1, \tag{II. 4}$$

where, for $S_0=A_0\oplus A_0^c$,

$$A_0\equiv(1,\tfrac{1}{2})\oplus(0,\tfrac{1}{2}), \tag{II. 5a}$$

$$A_0^c\equiv(\tfrac{1}{2},1)\oplus(\tfrac{1}{2},0) \tag{II. 5b}$$

and, for $S_1=A_1\oplus A_1^c$,

$$A_1=A_0\oplus(0,\tfrac{1}{2}), \tag{II. 6a}$$

$$A_1^c=A_0^c\oplus(\tfrac{1}{2},0). \tag{II. 6b}$$

(3) We assume that there exists an invariant, Hermitian bilinear form realized by a numerical matrix η satisfying

$$S^\dagger(\Lambda)\eta S(\Lambda)=\eta \quad \text{for all } \Lambda\in\mathbf{P}, \tag{II. 7a}$$

$$P\eta P^{-1}=\eta, \tag{II. 7b}$$

$$\eta=\eta^\dagger, \tag{II. 7c}$$

and

$$\eta\beta_\mu=\beta_\mu^\dagger\eta. \tag{II. 7d}$$

Note that we have *not* made the usual assumption that η is nonsingular.

(4) Equation (I. 1) is assumed to describe a particle of unique mass m. This means[6] that β_μ must satisfy the algebraic relation

$$(\beta\cdot p)^n=(\beta\cdot p)^{n-2}p^2 \tag{II. 8}$$

for some n where p is any 4-momentum. In terms of β_0 this becomes

$$\beta_0^{n-2}(\beta_0^2-1)=0. \tag{II. 9}$$

(5) Finally, we assume that Eq. (I. 1) describes a particle with a unique spin which we take for the cases considered to be $s=3/2$. This means that the solutions of Eq. (I. 1) when written in momentum space and taken to the rest frame will have only eight nonvanishing components corresponding to the four spin degrees of freedom for $s=3/2$ and the two signs of the energy.

The representations considered here contain only two spins, 3/2 and 1/2, the $s=3/2$ representation occurring twice and the $s=1/2$ representation occurring four and six times for the $S_0(\Lambda)$ and $S_1(\Lambda)$ representations, respectively. We may accordingly decompose β_0 into two submatrices which act only between components of the same spin:

$$\beta_0=[\beta_0]_{3/2}\oplus[\beta_0]_{1/2} \tag{II. 10}$$

and from assumption (5) we have the characteristic equations for these submatrices

$$[\beta_0]_{3/2}^2-1=0 \tag{II. 11a}$$

and

$$[\beta_0]_{1/2}^N=0 \tag{II. 11b}$$

for some N. From this viewpoint we see that it is the degree of nilpotency of the auxiliary spins which will determine the order of the algebra (II. 8).[7]

In the next two sections we shall seek the most general wave equations associated with the particular representations $S_0(\Lambda)$ and $S_1(\Lambda)$ when subject to the above restrictions.

III. THE STRUCTURE OF WAVE EQUATIONS TRANSFORMING VIA $S_0(\Lambda)$

Using the generators of $S_0(\Lambda)$[8] and invoking Eqs. (II. 1) and (II. 3a) we may write the general form of β_0 in this representation as

$$\beta_0=\begin{array}{c|cc} & A_0 & A_0^c \\ \hline A_0 & 0 & M \\ A_0^c & M & 0 \end{array} \tag{III. 1}$$

where

$$M=\begin{array}{c|ccc} & 3/2 & 1/2 & 1/2 \\ \hline 3/2 & m_{11} & 0 & 0 \\ 1/2 & 0 & -\tfrac{1}{2}m_{11} & m_{23} \\ 1/2 & 0 & m_{32} & m_{33} \end{array} \quad \begin{array}{l} \\ (1,\tfrac{1}{2}) \\ \\ (0,\tfrac{1}{2}) \end{array}$$

$$=\begin{array}{c|cc} & M_{3/2} & 0 \\ \hline M_{3/2} & 0 & \\ 0 & & M_{1/2} \end{array}, \qquad m_{ij}\in C, \tag{III. 2}$$

which satisfies assumptions (1) and (2).

Let us now invoke assumptions (4) and (5). We have from Eqs. (II. 11) that

$$M_{3/2}^2=m_{11}^2I=I \tag{III. 3}$$

and

$$M_{1/2}^N=0 \quad \text{for some } N. \tag{III. 4}$$

Equation (III. 3) implies that $m_{11}^2=1$ and we take $m_{11}=+1$. This involves no loss of generality and corresponds to the freedom of picking the overall sign of β_μ. Equation (III. 4) then implies that

$$\begin{array}{|cc|} \hline -1/2 & m_{23} \\ m_{32} & m_{33} \\ \hline \end{array}^N =0 \tag{III. 5}$$

for some N.

Since the cases $N=0$ and $N=1$ are not possible and

since $\dim(M_{1/2})=2$, $N=2$ is the only possible exponent in Eq. (III. 5). This leads to the conditions

$$m_{33}=\tfrac{1}{2} \tag{III. 6a}$$

and

$$m_{23}m_{32}=-\tfrac{1}{4}. \tag{III. 6b}$$

We shall now consider assumption (3). The most general η matrix for $S_0(\Lambda)$ satisfying Eqs. (II. 7a)—(II. 7c) may be written as

$$\eta=\begin{array}{c|cc} & A_0 & A_0^c \\ \hline A_0 & 0 & \hat{\eta} \\ A_0^c & \hat{\eta} & 0 \end{array} \tag{III. 7}$$

with

$$\hat{\eta}=\begin{array}{c|ccc} & 3/2 & 1/2 & 1/2 \\ \hline 3/2 & \eta_1 & 0 & 0 \\ 1/2 & 0 & \eta_1 & 0 \\ 1/2 & 0 & 0 & \eta_2 \end{array}, \quad \eta_i\in R. \tag{III. 8}$$

Equation (II. 7d) tells us that

$$\hat{\eta}M=M^{\dagger}\hat{\eta} \tag{III. 9}$$

or that m_{11} and m_{33} are real and

$$\eta_1\bar{m}_{23}=m_{32}\eta_2. \tag{III. 10}$$

Note that in view of Eq. (III. 6b) we must have η_1 and η_2 nonzero in order to avoid a trivial theory. We may also fix the sign and normalization of the rest frame solutions by taking $\eta_1=1$.

The most general β_0 satisfying assumptions (1)—(5) for $S_0(\Lambda)$ is thus found from Eqs. (III. 6) and (III. 9) and is given by Eq. (III. 1) with

$$M=\begin{array}{|c|c|c|}\hline 1 & 0 & 0 \\ \hline 0 & -\tfrac{1}{2} & \alpha \\ \hline 0 & -\bar{\alpha} & \tfrac{1}{2} \\ \hline \end{array} \tag{III. 11}$$

where $|\alpha|^2=\frac{1}{4}$. The elements of η in Eq. (III. 8) are $\eta_1=1$, $\eta_2=-1$.

Subject to the above assumptions there is therefore only one algebraic structure possible for $S_0(\Lambda)$, namely,

$$(\beta\cdot p)^2[(\beta\cdot p)^2-1]=0. \tag{III. 12}$$

This is the algebra which characterizes the R—S system.[9]

IV. THE STRUCTURE OF WAVE EQUATIONS TRANSFORMING VIA $S_1(\Lambda)$

Again by constructing the generators of $S_1(\Lambda)$ (Ref. 8) and imposing assumptions (1) and (2), we are led to the general form

$$\beta_0=\begin{array}{c|cc} & A_1 & A_1^c \\ \hline A_1 & 0 & N \\ A_1^c & N & 0 \end{array} \tag{IV. 1}$$

where

$$N=\begin{array}{|c|c|}\hline N_{3/2} & 0 \\ \hline 0 & N_{1/2} \\ \hline \end{array}, \tag{IV. 2}$$

$N_{3/2}$ is a multiple of the (four-dimensional) identity, $N_{3/2}=\lambda I$, and

$$N_{1/2}=\begin{array}{|c|c|c|}\hline -\tfrac{1}{2}\lambda & a_{12} & a_{13} \\ \hline a_{21} & a_{22} & a_{23} \\ \hline a_{31} & a_{32} & a_{33} \\ \hline \end{array}, \quad a_{ij}\in C. \tag{IV. 3}$$

The assumptions (4) and (5) will again be satisfied if and only if $\lambda^2=1$ (we again take $\lambda=+1$ with no loss of generality) and either

$$N_{1/2}^2=0, \quad N_{1/2}\neq 0 \tag{IV. 4}$$

or

$$N_{1/2}^3=0, \quad N_{1/2}^2\neq 0, \tag{IV. 5}$$

the degree being bounded by the dimensionality of the matrix. There are thus two distinct algebraic possibilities.

Consider now the most general η matrix satisfying Eqs. (II. 7a)—(11. 7c):

$$\eta=\begin{array}{c|cc} & A_1 & A_1^c \\ \hline A_1 & 0 & \hat{\eta} \\ A_1^c & \hat{\eta} & 0 \end{array} \tag{IV. 6}$$

with

$$\hat{\eta}=\begin{array}{c|cccc} & 3/2 & 1/2 & 1/2 & 1/2 \\ \hline 3/2 & \eta_1 & & & \\ 1/2 & & \eta_1 & & \\ 1/2 & & & \eta_2 & \eta_3 \\ 1/2 & & & \bar{\eta}_3 & \eta_4 \end{array}. \tag{IV. 7}$$

We may again fix the normalization of the $(1,\frac{1}{2})$ and $(\frac{1}{2},1)$ components such that $\eta_1=1$ and since $\hat{\eta}$ is Hermitian we may diagonalize it by taking a suitable linear combination of the identical $SL(2,C)$ representations, thus setting $\eta_3=0$ with no loss of generality. Finally, we may set the scale of the lower components such that we have

$$\hat{\eta}=\begin{array}{c|cccc} & 3/2 & 1/2 & 1/2 & 1/2 \\ \hline 3/2 & 1 & & & \\ 1/2 & & 1 & & \\ 1/2 & & & \mu & \\ 1/2 & & & & \nu \end{array} = \begin{array}{|c|c|}\hline \hat{\eta}_{3/2} & \\ \hline & \hat{\eta}_{1/2} \\ \hline \end{array} \tag{IV. 8}$$

where $\mu, \nu=+1, -1$ or 0.

Equation (II. 7d) will be satisfied for β_0 if and only if

$$\hat{\eta}_{1/2}N_{1/2}=N_{1/2}^{\dagger}\hat{\eta}_{1/2} \tag{IV. 9}$$

where the matrices are given by Eqs. (IV. 3) and (IV. 8). The conditions on the elements of $N_{1/2}$ are thus

$$\mu a_{21}=\bar{a}_{12}, \tag{IV. 10a}$$

$$\nu a_{31}=\bar{a}_{13}, \tag{IV. 10b}$$

$$\nu a_{32}=\mu\bar{a}_{23}, \tag{IV. 10c}$$

and a_{22} and a_{33} must be real.

We shall label the various cases according to the various possible values of (μ, ν). Noting that equivalent systems result if μ and ν are interchanged, there are six possibilities: (i) (1, 1), (ii) (- 1, - 1), (iii) (+ 1, - 1), (iv) (1, 0), (v) (- 1, 0), (iv) (0, 0).

(i) $(\mu, \nu)=(1, 1)$. Here Eqs. (IV. 10) imply that $N_{1/2}$ is Hermitian and hence diagonalizable by a unitary transformation which preserves $\hat{\eta}=I$. But Eqs. (IV. 4) or (IV. 5) imply that $N_{1/2}$ is nilpotent and so its eigenvalues must be zero. This implies that $N_{1/2}=0$ which contradicts Eq. (IV. 3) if we are to have a nontrivial theory. Case (i) must therefore be abandoned.

(ii) $(\mu, \nu)=(-1, -1)$. Here Eqs. (IV. 10) yield

$$N_{1/2}=\begin{array}{|c|c|c|}\hline -\frac{1}{2} & a & b \\ \hline -\bar{a} & c & e \\ \hline -\bar{b} & \bar{e} & d \\ \hline\end{array}\ . \tag{IV. 11}$$

In this case e may be set equal to zero by a suitable $\hat{\eta}$-preserving unitary transformation which diagonalizes $N_{1/2}$ on the lower two components.

(iia) If $N_{1/2}^2=0$ then it is easy to verify that either $a=0=c$ or $b=0=d$ and we are left with the R—S system of Sec. III trivially extended by zero entries in the additional components.

(iib) On the other hand, Eq. (IV. 5) may be satisfied if and only if

$$\det(N_{1/2}-\lambda)=-\lambda^3. \tag{IV. 12}$$

This places the restriction on the elements of $N_{1/2}$;

$$c+d-\tfrac{1}{2}=0, \tag{IV. 13a}$$

$$\tfrac{1}{4}-cd-|a|^2-|b|^2=0, \tag{IV. 13b}$$

and

$$(-\tfrac{1}{2}c+|a|^2)d+c|b|^2=0. \tag{IV. 13c}$$

If we choose any of the parameters in Eqs. (IV. 13) to be zero, then we are reduced to the case (iia) and a trivial extension of the R—S equation. Avoiding these cases, we may solve Eqs. (IV. 13) and find

$$d=\tfrac{1}{2}-c, \tag{IV. 14a}$$

$$|a|^2=c^3/(2c-\tfrac{1}{2}) \tag{IV. 14b}$$

$$|b|^2=(c-\tfrac{1}{2})^3/(2c-\tfrac{1}{2}), \tag{IV. 14c}$$

thus parameterizing the solutions in terms of c. We see that consistent nontrivial solutions exist whenever $c\in\langle-\infty,\ 0\rangle\cup\langle\frac{1}{2},\ \infty\rangle$. For the special choice $c=-\frac{1}{2}$ we get

$$N_{1/2}=\begin{array}{|c|c|c|}\hline -\frac{1}{2} & \frac{1}{2\sqrt{3}} & \frac{\sqrt{2}}{\sqrt{3}} \\ \hline \frac{-1}{2\sqrt{3}} & -\frac{1}{2} & 0 \\ \hline -\frac{\sqrt{2}}{\sqrt{3}} & 0 & 1 \\ \hline\end{array} \tag{IV. 15}$$

which reproduces the β-matrices presented by Glass[4,5] in his counterexample to the relation of Umezawa and Visconti.[10] We see how the order of the minimal equation for β_0 is increased from the $\beta_0^2(\beta_0^2-1)=0$ of the R—S equation to $\beta_0^3(\beta_0^2-1)=0$ of the Glass equation without increasing the highest value of the spin. We may also note that any choice of c which leads to nonzero a and b must yield a system which satisfies the higher order algebra.

(iii) $(\mu, \nu)=(1, -1)$.

$$N_{1/2}=\begin{array}{|c|c|c|}\hline -\frac{1}{2} & a & b \\ \hline \bar{a} & c & e \\ \hline -\bar{b} & -\bar{e} & d \\ \hline\end{array}\ , \tag{IV. 16}$$

$$\hat{\eta}=\begin{array}{|c|c|c|}\hline 1 & 0 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & 0 & -1 \\ \hline\end{array}\ . \tag{IV. 17}$$

The metric operator $\hat{\eta}$ will be preserved by any transformation of the form

$$T(\theta)=\begin{array}{|c|c|c|}\hline 1 & 0 & 0 \\ \hline 0 & \cosh\theta & \sinh\theta \\ \hline 0 & \sinh\theta & \cosh\theta \\ \hline\end{array}\ , \quad \theta\in R, \tag{IV. 18}$$

since $T^{*}\hat{\eta}T=\hat{\eta}$. If we first transform the $N_{1/2}$ of Eq. (IV. 16) by the unitary transformation

$$F=\begin{array}{|c|c|c|}\hline e^{i\alpha} & & \\ \hline & e^{i\beta} & \\ \hline & & e^{i\gamma} \\ \hline\end{array}\ , \tag{IV. 19}$$

then we may fix α, β, and γ such that all of the elements of $N_{1/2}$ are real. We now transform $N_{1/2}$ by $T(\theta)$:

$$N'_{1/2}=T(\theta)N_{1/2}T^{-1}(\theta)=\begin{array}{|c|c|c|}\hline -\frac{1}{2} & a' & b' \\ \hline a' & c' & e' \\ \hline -b' & -e' & d' \\ \hline\end{array} \tag{IV. 20}$$

where

$$a'=a\cosh\theta-b\sinh\theta,$$

$$b'=-a\sinh\theta+b\cosh\theta,$$

$$c'=c\cosh^2\theta-d\sinh^2\theta-2e\cosh\theta\sinh\theta,$$

$$d' = -c\sinh^2\theta + d\cosh^2\theta + 2e\cosh\theta\sinh\theta,$$

and

$$e' = e(\cosh^2\theta + \sinh^2\theta) + (d-c)\cosh\theta\sinh\theta.$$

We may now consider the following cases,

(a) If $|b| > |a|$ then we may have $a' = 0$ by choosing $\coth\theta = b/a$.

(b) If $|a| > |b|$ then we may have $b' = 0$ by choosing $\coth\theta = a/b$.

(c) $|a| = |b|$.

Now $\det(N_{1/2} - \lambda) = -\lambda^3$ will hold if and only if

$$c + d - \tfrac{1}{2} = 0, \tag{IV.21a}$$

$$\tfrac{1}{4} - cd - e^2 + a^2 - b^2 = 0, \tag{IV.21b}$$

$$-\tfrac{1}{2}cd - \tfrac{1}{2}e^2 - da^2 + cb^2 - 2abe = 0. \tag{IV.21c}$$

(a) Choose $a = 0$. Then Eqs. (IV.21) admit the solutions

$$d = \tfrac{1}{2} - c, \tag{IV.22a}$$

$$b^2 = \tfrac{1}{8}[1/(c+\tfrac{1}{2})], \tag{IV.22b}$$

$$e^2 = c^3/(c+\tfrac{1}{2}) \tag{IV.22c}$$

which are consistent for any $c \geq 0$.

(b) Choose $b = 0$. Then we get

$$d = \tfrac{1}{2} - c \tag{IV.23a}$$

$$a^2 = \tfrac{1}{8}[1/(c-1)], \tag{IV.23b}$$

$$e^2 = (2c-1)^3/8(c-1) \tag{IV.23c}$$

which are consistent for any $c > 1$.

From Eqs. (IV.22) we see that if $c = 0$ then $e = 0$ and we recover the trivially extended R–S equation and $N^2_{1/2} = 0$. For the remaining values of c we have $N^3_{1/2} = 0$ but $N^2_{1/2} \neq 0$.

(c) Choose $b = \epsilon a$, $\epsilon = \pm 1$. We then get from Eqs. (IV.21)

$$d = \tfrac{1}{2} - c, \tag{IV.24a}$$

$$e^2 = c^2 - \tfrac{1}{2}c + \tfrac{1}{4}, \tag{IV.24b}$$

$$a^2 = \tfrac{1}{8}(2c - \tfrac{1}{2} - 2\epsilon|e|)^{-1}. \tag{IV.24c}$$

If $\epsilon = -1$ then these conditions are consistent for all c, but if $\epsilon = +1$ they imply that $a^2 < 0$ for any c. One solution would be $c = 0$:

$$N_{1/2} = \tfrac{1}{2}\begin{array}{|c|c|c|}\hline -1 & 1 & -1 \\ \hline 1 & 0 & 1 \\ \hline 1 & -1 & 1 \\ \hline\end{array} \tag{IV.25}$$

which satisfies $N^3_{1/2} = 0$, $N^2_{1/2} \neq 0$. This is, in fact, again always the case for type (iii) matrices.

We now depart slightly from orthodoxy and consider singular η matrices. These will lead to new types of wave equations called "barnacled" equations.

(iv) and (v), $(\mu, \nu) = (\epsilon, 0)$, $\epsilon = \pm 1$. Here we have

$$N_{1/2} = \begin{array}{|c|c|c|}\hline -\tfrac{1}{2} & a & 0 \\ \hline \epsilon\bar{a} & c & 0 \\ \hline b & e & d \\ \hline\end{array} \tag{IV.26}$$

with $d = 0$ if we demand that $N_{1/2}$ be nilpotent. $\det(N_{1/2} - \lambda) = -\lambda^3$ implies that $c = \frac{1}{2}$ and $|a|^2 = -\epsilon\frac{1}{4}$. Thus case (iv) ($\epsilon = +1$) cannot be satisfied but case (v) does admit the solution

$$N_{1/2} = \begin{array}{|c|c|c|}\hline -\tfrac{1}{2} & \pm\tfrac{1}{2} & 0 \\ \hline \mp\tfrac{1}{2} & \tfrac{1}{2} & 0 \\ \hline b & e & 0 \\ \hline\end{array} \tag{IV.27}$$

which satisfies $N^2_{1/2} = 0$ only if $b = \mp e$. Otherwise, we get $N^3_{1/2} = 0$, $N^2_{1/2} \neq 0$.

(vi) $(\mu, \nu) = (0, 0)$. Here

$$N_{1/2} = \begin{array}{|c|c|c|}\hline -\tfrac{1}{2} & 0 & 0 \\ \hline a & c & e \\ \hline b & e' & d \\ \hline\end{array} \tag{IV.28}$$

which can never be nilpotent and hence this case must be abandoned.

Thus there are three classes of equations which satisfy all of our requirements. These correspond to η matrices of the classes (ii), (iii), and (iv). For the case of nonsingular η[(ii) and (iii)] the formalism either reduces to the R–S equation trivially extended by zeroes or increases by one the order of the β-algebra by increasing the degree of nilpotency of the spin-$\frac{1}{2}$ submatrix. As we shall see in the next section for the case of singular η, the "barnacled" equation is completely equivalent to the R–S equation in that its solutions are identical but the order of the β_0's minimal equation may be increased.

V. SUMMARY AND DISCUSSION

We have examined all possible wave equations subject to general physical assumptions which transform according to a given representation of $SL(2, C)$ for two special choices of representation. The first choice, $S_0(\Lambda)$, led to a unique algebraic structure

$$(\beta\cdot p)^2[(\beta\cdot p)^2 - p^2] = 0 \tag{V.1}$$

and consequently to the R–S equation. The second choice, $S_1(\Lambda)$, either (1) collapsed to a trivial extension of the former choice obeying the same algebra or (2) led to an algebra which is of higher degree than Eq. (V.1):

$$(\beta\cdot p)^3[(\beta\cdot p)^2 - 1] = 0. \tag{V.2}$$

One such case reproduced the theory presented by Glass. Higher algebras were seen to be prohibited by the dimensions of the nilpotent submatrices involved.

Finally, we considered a class of equations which admit only a singular η-matrix. Let us consider this new class in more detail.

We have from Eq. (IV. 27) (indicating the dimension of the submatrices)

$$\beta_0 = \begin{array}{c|c|c} & \begin{matrix}10 & 10\end{matrix} & \\ \hline \begin{matrix}0\\N\end{matrix} & \begin{matrix}N\\0\end{matrix} & \begin{matrix}10\\10\end{matrix} \end{array}, \quad N = \begin{array}{c|c|c} & \begin{matrix}4 & 6\end{matrix} & \\ \hline \begin{matrix}N_{3/2}\\0\end{matrix} & \begin{matrix}0\\N_{1/2}\end{matrix} & \begin{matrix}4\\6\end{matrix} \end{array}, \tag{V.3}$$

and

$$N_{1/2} = \begin{array}{c|c|c|c} \begin{matrix}2 & 2 & 2\end{matrix} & & & \\ \hline -\frac{1}{2} & \frac{1}{2} & 0 & 2 \\ \hline -\frac{1}{2} & \frac{1}{2} & 0 & 2 \\ \hline b & e & 0 & 2 \end{array}. \tag{V.4}$$

The matrices thus take the structure (diagonalizing P)

$$\beta_0 = \begin{array}{c|c|c} \begin{matrix}10 & 10\end{matrix} & & \\ \hline N & & 10 \\ \hline & -N & 10 \end{array} \tag{V.5}$$

which may be written in the form

$$\beta_0 = \begin{array}{c|c|c} \begin{matrix}16 & 4\end{matrix} & & \\ \hline \Gamma_0 & 0 & 16 \\ \hline B_0 & 0 & 4 \end{array} \tag{V.6}$$

where Γ_0 is the matrix for the R—S equation and B_0 represents the arbitrary coupling of the extra 4-spinor components. The wave equation thus takes the form

$$\left\{ i \begin{array}{c|c} \Gamma_\mu & 0 \\ \hline B_\mu & 0 \end{array} \partial^\mu - m \right\} \begin{pmatrix} \psi \\ \psi_B \end{pmatrix} = 0 \tag{V.7}$$

from which we recover the R—S equation

$$(i\Gamma \cdot \partial - m)\psi = 0 \tag{V.8}$$

for the 16 upper components and

$$iB \cdot \partial \psi = m\psi_B \tag{V.9}$$

for the four lower components. Once an R—S solution ψ is given, then ψ_B is completely determined. Hence the solutions to Eq. (V. 7) are the same as those of Eq. (V. 8). Furthermore, the components ψ_B do not enter the scalar product because the η matrix now acts as a projector onto the ψ components. Thus the system (V. 7) is completely equivalent to the R—S system. Hence we have called the additional components ψ_B which transform like a Dirac 4-spinor "barnacles."

Note, however, that although the barnacled equation defines the same set of solutions as the unbarnacled equation, the minimal equation of the β_0 matrix may be changed for certain choices of the barnacle coupling. Thus we have two equations with two different algebras but with the same solutions.

It is easy to see from Eq. (V. 7) that all of the above remarks remain valid even in the presence of a minimally coupled external electromagnetic field, $\partial_\mu \to \partial_\mu + ieA_\mu$. Indeed we see that a more general external field coupling formed from any product of β-matrices[5,11] could never involve the ψ_B components in the interaction, i.e., the equation for ψ is unchanged by the presence of ψ_B. Barnacles may, however, enter the dynamics in a nontrivial way for other couplings which are now possible since the β_μ are reducible (but indecomposable). They may therefore provide a convenient vehicle for modifying the dynamics of a physical process while being guaranteed to be absent in the asymptotic states which are governed by the free equation of motion. A more general and detailed discussion will be presented in a subsequent study.

We have seen that simple algebraic techniques may be applied to a given representation of $SL(2, C)$ in order to find the structure of all possible relativistic wave equations corresponding to that representation. This procedure may, of course, be applied to any representation. The present cases merely served (we hope) as illustrative examples.

*Supported by the U. S. Atomic Energy Commission AT(40-1) 3992.

[1]W. J. Hurley and E. C. G. Sudarshan, Ann. Phys. (N.Y.) 85, 111 (1974).

[2]The study of the algebraic structure of first order wave equations has a long history. Some early references are N. Kemmer, Proc. Roy. Soc. A 173, 91 (1939) and H. J. Bhabha, Rev. Mod. Phys. 17, 200 (1945); 21, 451 (1949); in addition, to the work cited in Ref. 5. For a review and a general discussion see the books: E. M. Corson, *Introduction to Tensors, Spinors and Relativistic Wave Equations* (Blackie, London, 1953); I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (Pergamon, New York, 1963); and M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon, New York, 1964). Some recent applications may be found in A. Z. Capri, Phys. Rev. 178, 2427 (1969) and A. Shamaly and A. Z. Capri, Nuovo Cimento B 2, 236 (1971).

[3]W. Rarita and J. Schwinger, Phys. Rev. 60, 61 (1941); S. N. Gupta, Phys. Rev. 95, 1334 (1954).

[4]A. S. Glass, Commun. Math. Phys. 23, 176 (1971).

[5]A. S. Glass, Ph. D. thesis, Princeton University, 1971 (unpublished).

[6]J. Weinberg, Ph. D. thesis, University of California, 1943 (unpublished); Harish-Chandra, Phys. Rev. 71, 793 (1947).

[7]That the role of subsidiary conditions is to eliminate the unwanted auxiliary spins in a given theory has been emphasized, e.g., by D. L. Pursey, Ann. Phys. (N.Y.) 32, 157 (1965) in his general study. Here we have first order wave equations without subsidiary conditions and we see that the elimination of the unwanted spins is accomplished by the nilpotency of β_0 on the given subspace thus guaranteeing that the auxiliary spins will vanish in the rest frame.

[8]See Ref. 1, Appendices A and B.

[9]For another approach to this classification see K. Johnson and E. C. G. Sudarshan, Ann. Phys. (N.Y.) 13, 126 (1961).

[10]H. Umezawa and A. Visconti, Nucl. Phys. 1, 348 (1956).

[11]A. S. Wightman, in *Aspects of Quantum Theory*, edited by A. Salam and E. P. Wigner (Cambridge U. P., Cambridge, 1972), pp. 95—115; and Proc. Symp. Pure Math. 23, 441 (1972).

On some geometrical aspects of classical nonconservative mechanics

Dj. Djukic and B. Vujanovic

Faculty of Technical Sciences, University of Novi Sad, 21000 Novi Sad, Yugoslavia
(Received 31 December 1974)

It is shown that a scleronomous, holonomic dynamical system with nonconservative forces moves in such a way that the differential equations of motion are geodesic lines in a linear connected space L_n. The space L_n is semimetric and semisymmetric. The geodesic line on which the tangent at a point remains tangent if it is parallel displaced along the curve is simultaneously the curve of stationary length between two points in the space L_n. A necessary condition for the stationary length is derived by making use of the noncommutation rule for the differential of variation and the variation of differential. The noncommutation rule is obtained from a quadrilateral, which is called the fundamental quadrilateral of variational calculus. By using the noncommutation rule, the variational principles of Maupertius and Hamiltonian type for nonconservative mechanical systems are presented.

1. INTRODUCTION

Under the concept of geometrization of motion we usually mean finding the space in which the differential equations of motion are identical with geodesic lines of the space. After the classical papers[1-3] there were many attempts to describe the motion of nonconservative dynamical systems with the help of non-Riemannian geometry. It was shown[4-8] that nonconservative mechanical systems imply spaces with a more complex structure than the Riemannian. In the metrical sense they are usually Weyl's semimetrical spaces. The connection coefficients of the spaces contain together with the Christoffel symbols, formed with respect to the fundamental metric tensor of the space, a part which is dependent on the nonconservative forces. The geodesic lines in this space are the curves for which the tangent vector at a point remains tangent if it is parallel displaced along the curve. In the space where the connection coefficients are calculated by use of only the fundamental metric tensor of the space, these geodesic lines defined by the parallel displacement are simultaneously curves of extremal length between two points in the space (Reimannian space, Finsler space, etc.). In the spaces of nonconservative mechanical systems this unique meaning of the geodesic lines is not preserved.

It is obvious that if we want to maintain this elegant coincidence in the case of nonconservative mechanical systems, we must make some fundamental changes in the classical variational calculus, or, what amounts to the same thing, some changes in the quality of the functional's extremum. Recent papers[10,11] provide a possibility for this.

An linear connected space L_n is constructed, whose geodesic lines are equivalent to the differential equations of motion of the nonconservative mechanical system. The space L_n is semimetric and semisymmetric. The geodesic lines are simultaneously curves of stationary length between two points in the space L_n. The stationarity condition is obtained using the noncommutation rule[10,11] for the differential of variation and the variation of differential. Using the same technique the variational principles of Maupertius and Hamiltonian type for nonconservative mechanical system are presented.

2. A GEOMETRIZATION OF MOTION

Let us consider a holonomic scleronomic nonconservative mechanical system with n degrees of freedom, where the q^i are regarded as the generalized coordinates and t is the time. The kinetic energy of the system is

$$T=\tfrac{1}{2}g_{\lambda\mu}\dot{q}^{\lambda}\dot{q}^{\mu}, \tag{1}$$

where $g_{\lambda\mu}$ is the fundamental covariant metric tensor of second order which is a function of position q^{λ}, Greek indices imply a range of values from 1 to n, and, when repeated in a single term, summation over that range. Let the system be subject to the nonconservative generalized forces Q_{μ}, which are functions of the generalized coordinates, the generalized velocities $\dot{q}^{\lambda}$, and the time, i.e., $Q_{\mu}=Q_{\mu}\ (t,q,\dot{q})$. Then the governing equations of motion in contravariant form are

$$\frac{\overset{\varepsilon}{\delta}\dot{q}^{\omega}}{\delta t}\equiv\ddot{q}^{\omega}+\begin{Bmatrix}\omega\\ \lambda\ \mu\end{Bmatrix}\dot{q}^{\lambda}\dot{q}^{\mu}=Q^{\omega}\quad(\cdot\equiv d/dt), \tag{2}$$

where $\{\lambda^{\omega}{}_{\mu}\}$ are Christoffel symbols formed with respect to the tensor $g_{\lambda\mu}$, and the symbol $\overset{\varepsilon}{\delta}/\delta t$ designates the absolute derivative with respect to the same tensor $g_{\lambda\mu}$. By a simple direct calculation we can prove that the following theorem is valid:

Theorem I: A scleronomic, holonomic dynamical system, with nonconservative forces Q_{μ} which are functions of position, velocities, and time, moves in such a way that the differential equations of motion (2) are identical with the geodesic lines

$$\frac{\overset{\Gamma}{\delta}\dot{q}^{\omega}}{\delta t}\equiv\ddot{q}^{\omega}+\Gamma^{\omega}_{\lambda\mu}\dot{q}^{\lambda}\dot{q}^{\mu}=0, \tag{3}$$

in the linear connected space L_n with connection coefficients

$$\Gamma^{\omega}_{\lambda\mu}=\begin{Bmatrix}\omega\\ \lambda\ \ \mu\end{Bmatrix}+\frac{1}{2T}(-Q_\lambda\delta^\omega_\mu+Q_\mu\delta^\omega_\lambda-Q^\omega g_{\lambda\mu}),\tag{4}$$

where δ^ω_μ are the Kronecker deltas and $\overset{\Gamma}{\delta}/\delta t$ is the absolute derivative with respect to the connection coefficients $\Gamma^\omega_{\lambda\mu}$. During the motion, the time t plays the role of an affine parameter.

The connection coefficients can be written in the form

$$\Gamma^{\omega}_{\lambda\mu}=\Gamma^{\omega}_{(\lambda\mu)}+\Gamma^{\omega}_{[\lambda\mu]},\tag{5}$$

where

$$\Gamma^{\omega}_{(\lambda\mu)}=\begin{Bmatrix}\omega\\ \lambda\ \ \mu\end{Bmatrix}-\frac{1}{2T}Q^\omega g_{\lambda\mu}\tag{6}$$

is their symmetric part with respect to the covariant indices, while the antisymmetric part, or the torsion tensor (Ref. 13, p. 126) of the space L_n, is

$$\Gamma_{[\lambda}{}^{\omega}{}_{\mu]}=S_{[\lambda}\delta^{\omega}{}_{\mu]}\tag{7}$$

where the vectors S_λ are

$$S_\lambda=-(1/T)Q_\lambda.\tag{8}$$

The covariant derivative $(\overset{\Gamma}{\nabla})$ of the fundamental tensor $q_{\lambda\mu}$ with respect to the connection coefficients (4) is given by

$$\overset{\Gamma}{\nabla}_\nu g_{\lambda\mu}=\frac{1}{T}Q_\nu g_{\lambda\mu}.\tag{9}$$

The expressions for the covariant derivative and for the connection coefficients (4), together with Eqs. (5)–(8), determine the linear connected space L_n as a semimetric and semisymmetric space (see Ref. 13, p. 126). In the metrical sense this space is of Weyl's type. But the presence of the torsion tensor in the connection coefficients makes this space L_n more complicated than Weyl's space.

3. A FUNDAMENTAL QUADRILATERAL OF VARIATIONAL CALCULUS

The motion of a scleronomic holonomic nonconservative mechanical system can be considered as the motion of a representative point in the configuration space L_n. Let Γ and Γ^* (Fig. 1) be two neighbouring paths. The Γ is a direct path of the mechanical system and Γ^* is a varied path. Let M_1 and M_2 be two points on the direct path Γ, which correspond to the instants of time t and $t+dt$. If $\mathbf{r}$ is the radius vector of the point M_1, we have that $\mathbf{M_1M_2}=d\mathbf{r}$ and the radius vector of the point M_2 is $\mathbf{r}+d\mathbf{r}$. Further, let M_1' and M_2' be the points on the varied path corresponding to the instants t and $t+dt$, respectively. We will define the variation of the radius vector as the vector $\mathbf{M_1M_1'}=\delta\mathbf{r}$. We will assume that the vector $\mathbf{M_1'M_2'}$ on the varied path Γ^* is $\mathbf{M_1'M_2}'=d\mathbf{r}'+d\delta\mathbf{r}$, where $d\mathbf{r}'$ is a vector obtained by parallel displacement of the vector $d\mathbf{r}$ from the point M_1 to the point M_1'. The variation of the vector $d\mathbf{r}$, i.e., the vector $\delta d\mathbf{r}$, will be defined in such a way that $\mathbf{M_2M_2'}=\delta\mathbf{r}+\delta d\mathbf{r}$. Now from Fig. 1 it is obvious that the following vector equation is valid:

$$\delta\mathbf{r}+d\mathbf{r}'+d\delta\mathbf{r}=d\mathbf{r}+\delta\mathbf{r}+\delta d\mathbf{r}.\tag{10}$$

Introducing a metric form in the space L_n by the equation

$$(dS)^2=2T(dt)^2=g_{\alpha\beta}dq^\alpha dq^\beta,\tag{11}$$

where dS is the length of the vector $d\mathbf{r}(dS=|d\mathbf{r}|)$, we can calculate the change of length, using (9), as

$$\overset{\Gamma}{\nabla}(dS)=\frac{dS}{2T}Q_\nu\delta q^\nu\tag{12}$$

when the vector $d\mathbf{r}$ is parallel displaced $[\overset{\Gamma}{\nabla}(dq^\alpha)=0]$ from the point M_1 to the point M_1'. Here dq^α_Γ and δq^α are components of the vectors $d\mathbf{r}$ and $\delta\mathbf{r}$, and $\overset{\Gamma}{\nabla}$ is the covariant differential with respect to the connection coefficients $\Gamma^\alpha_{\beta\gamma}$.

Now, we have the value of vector $d\mathbf{r}'$,

$$d\mathbf{r}'=d\mathbf{r}+(d\mathbf{r})_0\overset{\Gamma}{\nabla}(dS),\tag{13}$$

where $(d\mathbf{r})_0=d\mathbf{r}/dS$ is the unit vector of direction $d\mathbf{r}$. Combining (12), (13) we obtain

$$d\mathbf{r}'=\left(1+\frac{1}{2T}Q_\nu\delta q^\nu\right)d\mathbf{r}.\tag{14}$$

Substituting (14) into (10) and using the equation (see Ref. 14, p. 29)

$$d\delta\mathbf{r}-\delta d\mathbf{r}=\frac{\partial\mathbf{r}}{\partial q^\alpha}(d\delta q^\alpha-\delta\,dq^\alpha),\tag{15}$$

we have

$$\frac{\partial\mathbf{r}}{\partial q^\alpha}\left(d\delta q^\alpha-\delta dq^\alpha+\frac{1}{2T}Q_\nu\delta q^\nu\,dq^\alpha\right)=0,\tag{16}$$

and finally (see Ref. 14, p. 30)

$$\delta dq^\alpha=d\delta q^\alpha+\frac{1}{2T}dq^\alpha Q_\nu\delta q^\nu.\tag{17}$$

This equation shows that in the linear connected space L_n with connection coefficients (4), the variation operator δ, and the operator of differentiating d do not commute. This fact was postulated in Ref. 10. In Ref. 11 the noncommutation rule (17) was introduced in agreement with the *central Lagrangian equation* (see Ref. 14, p. 253), while, in the present paper, we have shown its geometrical background. The noncommutation rule (17) is a consequence of the quadrilateral M_1, M_1', M_2', M_2 and its construction. Therefore, the quadrilateral may be called the fundamental quadrilateral of variational

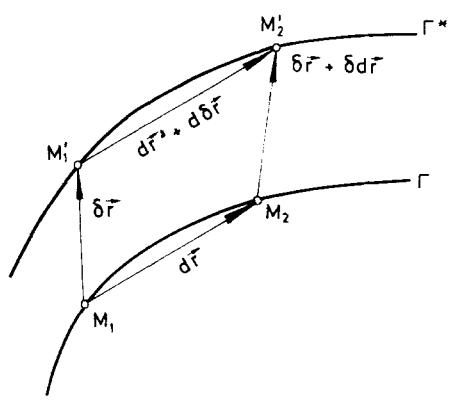


FIG. 1

calculus. When we have an inertial motion ($Q_\alpha \equiv 0$) of the mechanical system, then from (4) and (9) it is obvious that the space L_n becomes Riemannian. Under the same circumstances Eq. (17) yields the commutation rule for the operators δ and d.

4. GEODESIC LINES IN THE SPACE L_n AS THE CURVES OF EXTREMAL LENGTH

The extremal length of a curve joining two points A, B in the space L_n is given by the extremal value of the functional

$$S = \int_{t_0}^{t_1} \varphi(q^\lambda, \dot{q}^\lambda)\, dt, \tag{18}$$

where [see (11)]

$$\dot{S} \equiv \varphi(q^\lambda, \dot{q}^\lambda) = \sqrt{g_{\alpha\beta}\dot{q}^\alpha \dot{q}^\beta} \quad (2T = \dot{S}^2), \tag{19}$$

and where the instants t_0 and t_1 corresponds to the points A and B.

The first variation of (18) is

$$\delta S = \int_{t_0}^{t_1} \left(\frac{\partial \varphi}{\partial q^\alpha}\, \delta q^\alpha + \frac{\partial \varphi}{\partial \dot{q}^\alpha}\, \delta \dot{q}^\alpha \right) dt, \tag{20}$$

which by making use of (11), (17), and (19) and integration by parts becomes

$$\delta S = \frac{\partial \varphi}{\partial \dot{q}^\alpha}\, \delta q^\alpha \Big|_{t_0}^{t_1} + \int_{t_0}^{t_1} \left(\frac{\partial \varphi}{\partial q^\alpha} - \frac{d}{dt} \cdot \frac{\partial \varphi}{\partial \dot{q}^\alpha} + \frac{\dot{q}^\beta}{\dot{S}^2} Q_\alpha \frac{\partial \varphi}{\partial \dot{q}^\beta} \right) \delta q^\alpha\, dt. \tag{21}$$

If $\delta q^\alpha = 0$ for t_0 and t_1, then the extremal value of the functional (18) is obtained for $\delta S = 0$. This, together with the fact that the δq^α are independent variations, yields

$$\frac{\partial \varphi}{\partial q^\alpha} - \frac{d}{dt} \frac{\partial \varphi}{\partial \dot{q}^\alpha} + \frac{\partial \varphi}{\partial \dot{q}^\beta} \frac{\dot{q}^\beta}{\dot{S}^2} Q_\alpha = 0. \tag{22}$$

Substituting (19) into (22) gives

$$\frac{\overset{\epsilon}{\delta \dot{q}^\omega}}{\delta t} \equiv \ddot{q}^\omega + \left\{ \begin{matrix} \omega \\ \lambda \quad \mu \end{matrix} \right\} \dot{q}^\lambda \dot{q}^\mu = Q^\omega + \frac{\ddot{S}}{\dot{S}} \dot{q}^\omega, \tag{23}$$

and combining this result with (4) and (11) we have

$$\frac{\overset{\Gamma}{\delta \dot{q}^\omega}}{\delta t} \equiv \ddot{q}^\omega + \Gamma^\omega_{\lambda\mu} \dot{q}^\lambda \dot{q}^\mu = \frac{\ddot{S}}{\dot{S}} \dot{q}^\omega. \tag{24}$$

Selecting the variable parameter of the curve to be the arc of the curve, i.e., $t = S$, we see that these equations become Eqs. (3). Hence, we can formulate the following theorem:

Theorem II: The curve giving the extremal length of a curve between two points in the space L_n is simultaneously the curve on which the tangent vector at a point remains tangent if it is parallel displaced along the curve.

Hence, by this theorem, one unifying definition of the geodesic lines which holds in the Riemannian spaces is extended to the linear connected spaces L_n with connection coefficients (4).

Remark I: In this section we used the expression "extremal value of the functional," without any analysis of the second variation. Therefore, an expression "stationary value..." is much more appropriate.

Remark II: The stationary value of the functional (18) is obtained by the noncommutation rule (17). The rule may be considered as a restriction imposed on the family of curves which may be extremals. Hence, the problem belongs to the class of classical variational problems involving a conditional extremum.

Remark III: Theorem II is proved in the linear connected space L_n with connection coefficients (4). This implies that to every linear connected space which is different of the L_n there corresponds an equivalent noncommutation rule of the form (17).

5. A VARIATIONAL PRINCIPLE OF MAUPERTIUS' TYPE

Let us consider the stationary value of the functional

$$I = \int_{t_0}^{t_1} T(q^\alpha, \dot{q}^\alpha)\, dt, \tag{25}$$

where T is the kinetic energy (1), subject to the noncommutation rule (17). After a simple calculation we obtain

$$I = \frac{\partial T}{\partial \dot{q}^\alpha}\, \delta q^\alpha \Big|_{t_0}^{t_1} + \int_{t_0}^{t_1} \left(\frac{\partial T}{\partial q^\alpha} - \frac{d}{dt} \frac{\partial T}{\partial \dot{q}^\alpha} + Q_\alpha \right) \delta q^\alpha\, dt. \tag{26}$$

Now, the stationarity condition $\delta I = 0$, for $\delta q^\alpha |_{t_0}^{t_1} = 0$ and independent variations δq^α, furnishes the well-known equations of motion of a mechanical system with generalized forces Q_α.

This variational problem yields the functional (25) which is similar to the functional of the classical Maupertius variational principle. But the difference is obvious. Maupertius, principle requires time to be varied, while here this is not the case. Furthermore, Maupertius' principle is valid only for mechanical systems with constant energy. An essential similarity is important. The classical Maupertius principle in Jacobi's form (see Ref. 14, p. 713) gives the equations of motion as the geodesic lines in a Riemannian space with metric tensor $\bar{g}_{\alpha\nu} = (h - \Pi) g_{\alpha\nu}$, where Π is the potential energy and h is the constant total energy of the mechanical system. This idea is maintained here. The equations of motion which follow from the above variational principle are equivalent to Eqs. (2), i.e., to the equations of the geodesic lines (3) in the linear connected space L_n.

6. A VARIATIONAL PRINCIPLE OF HAMILTON'S TYPE[10,11]

The corresponding functional is

$$I = \int_{t_0}^{t_1} (T - \Pi)\, dt, \tag{27}$$

where the kinetic energy T has the form (1). If we generalize the noncommutation rule (17) to

$$\delta\, dq^\alpha = d\delta q^\alpha + \frac{dq^\alpha}{2T} Q^*_\nu\, \delta q^\nu, \tag{28}$$

where Q_ν^* are the generalized nonconservative forces, then, using the same technique as in the previous section, from (27) we obtain the corresponding differential equations of motion

$$\frac{d}{dt}\cdot\frac{\partial T}{\partial \dot{q}^\alpha}-\frac{\partial T}{\partial q^\alpha}=-\frac{\partial \Pi}{\partial q^\alpha}+Q_\alpha^*. \tag{29}$$

This variational principle of Hamilton's type can be easily generalized to rheonomic holonomic mechanical systems. The corresponding functional-action integral is again (27), but with the kinetic energy

$$T=T_0(t,q^\alpha)+T_1+T_2, \tag{30}$$

where

$$T_1=g_\lambda(t,q^\alpha)\dot{q}^\lambda,\quad T_2=\tfrac{1}{2}g_{\alpha\mu}(t,q^\lambda)\dot{q}^\alpha\dot{q}^\mu, \tag{31}$$

while the analog of (17) and (28) is

$$\delta\,dq^\alpha=d\delta q^\alpha+\frac{dq^\alpha}{2T_2+T_1}Q_\nu^*\delta q^\nu. \tag{32}$$

[1]J. L. Synge, Phil. Trans. Roy. Soc. Lond. A **226**, 31 (1926).
[2]J. L. Synge, *Tensorial Methods in Dynamics* (Toronto U. P., Toronto, 1936).
[3]J. Douglas, Ann. Math. **2**, 29 (1928).
[4]E. Clauser, Inst. di Mat. del Politec. di Milano **166**, 33 (1955).
[5]E. Clauser, Inst. di Mat. del Politec. di Milano **157**, 495 (1955).
[6]A. Lichnerowicz, C.R. Acad. Sci. **212**, 328 (1941).
[7]B. Vajanovic, Z.Angew. Math. Mech. **47**, 63 (1967).
[8]B. Vujanovic, Tensor, N.S. **21**, 48 (1970).
[9]J. Favard, *Cours de géométrie différentielle locale* (Gauthier-Villars, Paris, 1957).
[10]B. Vujanovic, Acta Mechanica **19**, 259 (1974).
[11]B. Vujanovic, "A Variational Principle For Non-Conservative Dynamical Systems," Z. Angew. Math. Mech. (to be published).
[12]Dj. Djukic and B. Vujanovic, "Noether's Theory in Classical Nonconservative Mechanics," Acta Mechanica (to be published).
[13]J.A. Schouten, *Ricci Calculus* (Springer-Verlag, Berlin, 1954).
[14]A.I. Lur'e, *Analytical Mechanics* (G.I.F.M.L., Moscow, 1961).

Convergence of the perturbative approach to the N-body problem in the $N\to\infty$ limit

F. Calogero

Istituto di Fisica, Università di Roma, 00185 Roma, Italy
and Istituto Nazionale di Fisica Nucleare, Sezione di Roma, Roma, Italy

F. Palumbo

Laboratori Nazionali del CNEN, 00044 Frascati, Italy
and Scuola di Perfezionamento in Fisica, Università di Roma, Roma, Italy
(Received 12 August 1974)

An N-body system with interparticle forces that are attractive at short range collapses in the limit $N\to\infty$, namely in this limit the ground-state energy per particle diverges to negative infinity. If instead the forces are sufficiently repulsive at short range, in the limit $N\to\infty$ saturation occurs, leading to a finite ground-state energy per particle ϵ. This quantity depends, among other things, on the "coupling constant" g (entering as a factor that multiplies the interaction), and the above remark clearly implies that it is defined in the $N\to\infty$ limit only for positive values of g (although, for finite N, it is defined both for positive and negative g). This fact is generally taken to imply that the function $\epsilon(g)$ is nonanalytic in g at $g=0$ and, therefore, that the perturbative expansion of $\epsilon(g)$, being a power expansion in g, is necessarily nonconvergent (although it might be asymptotic). The purpose of this paper is to demonstrate the lack of cogency of this argument. It is therefore concluded that nonconvergence of the perturbative expansion for $\epsilon(g)$ is thus far an unproven hypothesis. The lack of cogency of analogous current arguments concerning the equilibrium density of many-body systems is also pointed out.

Let us consider the N-body system characterized by the Hamiltonian

$$H=(2m)^{-1}\sum_{i=0}^{N}p_i^2+g\sum_{i>j=1}^{N}v(r_{ij}), \tag{1}$$

where of course $\mathbf{p}_i=-i\hbar\nabla_i$ and $r_{ij}=|\mathbf{r}_i-\mathbf{r}_j|$. Let $E(N;\rho,\ g)$ be the ground state energy of this system when confined within a container of volume $V=N/\rho$, namely the lowest eigenvalue of the equations

$$H\chi(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N)=E(N;\ \rho,\ g)\chi(\mathbf{r}_1,\ \mathbf{r}_2,\ldots,\mathbf{r}_N), \tag{2}$$

$$\chi(\mathbf{r}_1,\mathbf{r}_2\ldots,\mathbf{r}_N)=0 \quad \text{if } \mathbf{r}_1 \text{ or } \mathbf{r}_2 \text{ or } \cdots \mathbf{r}_N \text{ on surface } S \text{ of volume } V. \tag{3}$$

Under these conditions the quantity

$$\rho=N/V \tag{4}$$

represents the mean particle density of the system in the container; this quantity coincides with the actual density of the system only if this fills uniformly the whole container.

If the potential $v(r)$ is integrable at infinity and is sufficiently repulsive (positive) at short range, as we assume hereafter, for $g>0$ the system exhibits saturation,[1] namely the ("thermodynamic," i.e., macroscopic) limit

$$\epsilon(\rho,\ g)=\lim_{N\to\infty}[E(N;\ \rho,\ g)/N],\quad g>0, \tag{5}$$

exists and is finite. This limiting value represents the energy per particle in a macroscopic sample; it is certainly negative if the system is self-bound (as we assume hereafter[2]) and if its equilibrium (average) density $\bar{\rho}(g)$ exceeds the mean density in the container ρ, so that the (macroscopic) system occupies only part of the (macroscopic) container. Moreover, under such conditions, the energy per particle is clearly independent of ρ. If instead $\rho>\bar{\rho}(g)$, the system is compressed within the container, and therefore $\epsilon(\rho,\ g)$ does depend on ρ, being presumably an increasing function of ρ. Thus, as a function of ρ, $\epsilon(\rho,\ g)$ has the behavior indicated in Fig. 1.[3] The quantity $\bar{\epsilon}(g)$ (see Fig. 1) is then the energy per particle in a self-bound system of macroscopic size, and $\bar{\rho}(g)$ (see Fig. 1) is the corresponding density (or average density, if the equilibrium configuration is not homogeneous, as is for instance the case for crystals). A basic task of many-body theory is the evaluation of $\bar{\epsilon}(g)$ and $\bar{\rho}(g)$.

In principle the quantities $\bar{\epsilon}(g)$ and $\bar{\rho}(g)$ could be defined without any reference to a container; one should take the

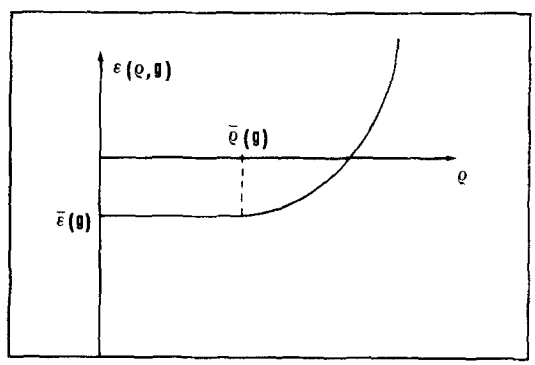


FIG. 1.

$N \to \infty$ limit in the problem characterized by the N-body Schrödinger equation (2), with no boundary conditions besides the asymptotic restrictions implied by the requirement that the many-body wavefunction χ be normalizable (in the center-of-mass frame). This is of course equivalent to the treatment outlined above, but with $\rho = 0$. There are, however, two reasons why the consideration of a container is useful, even though the corresponding problem is somewhat more complex, as indicated by the dependence of $\epsilon(\rho, g)$ upon ρ, that is clearly not analytic at $\rho = \bar{\rho}(g)$ (see Fig. 1).

The first reason is connected with the use of a perturbative approach; the presence of a container avoids the technical complication that would otherwise be associated with the continuous spectrum of the kinetic energy part of the Hamiltonian. Indeed, for regular potentials,[4] $E(N; \rho, g)$ is then an analytic function of g at $g = 0$, so that the power expansion

$$E(N; \rho, g) = \sum_{n=0}^{\infty} E_n(N; \rho) g^n \tag{6}$$

has a nonvanishing radius of convergence (for N and V finite). Thus, for sufficiently weak forces, perturbation theory is then convergent—although the radius of convergence might tend to zero as N tends to infinity (see below).

A second, and related, reason, has to do with the nonexistence, in the absence of a container, of any normalizable eigenstate of H for very small values of g; while to discuss the analyticity in g at $g = 0$ it is, of course, just on such values of g that attention is focused.[5]

For $g < 0$, the interaction $gv(r)$ becomes attractive at short range, and as a consequence the system, in the thermodynamic limit, collapses,[6] namely

$$[E(N; \rho, g)/N] \underset{N \to \infty}{\to} -\infty, \quad g < 0. \tag{7}$$

This fact is generally taken to imply that the function $\epsilon(\rho, g)$ of Eq. (5) is not analytic in g at $g = 0$, and therefore that any perturbative approach to the computation of $\epsilon(\rho, g)$ [and, *a fortiori*, $\bar{\epsilon}(g)$ and $\bar{\rho}(g)$] involves the handling of nonconvergent (although possibly asymptotic) expansions. The purpose of this communication is to point out that such conclusion lacks altogether cogency. We also note that, in spite of the nonanalytic ρ dependence of $\epsilon(\rho, g)$ displayed in Fig. 1, the functions $\bar{\epsilon}(g)$ and $\bar{\rho}(g)$, namely the quantities of physical interest, might well be analytic (and even entire), and so might be the function $\tilde{\epsilon}(\rho, g)$ defined to coincide with $\epsilon(\rho, g)$, Eq. (5), for $g > 0$ and $\rho > \bar{\rho}(g)$, and by analytic continuation for other values of g and ρ; of course, the values of $\tilde{\epsilon}(\rho, g)$ when g and ρ do not satisfy the above inequalities would then have no direct physical interpretation.

The simplest way to prove our point is to display a counterexample, namely to produce an explicit function $E(N; \rho, g)$ (artificially concocted), that does have the properties (5) [with $\epsilon(\rho, g)$ having the behavior shown in Fig. 1], (6), and (7), and yet yields function $\tilde{\epsilon}(\rho, g)$, $\bar{\epsilon}(g)$ and $\bar{\rho}(g)$ that are analytic (indeed entire). Having in mind a detailed recent analysis, in the framework of the many-fermion problem, of this question,[7] in which the conclusion that we consider unwarranted plays a crucial role, we require moreover that the function $E(N; \rho, g)$ have all the additional properties, that have been shown to hold in this case. This allows us to conclude, also in the context of the physically most interesting case, that nonconvergence of the perturbative expansion for the energy per particle, and the density, of the ground state of an N-body system in the $N \to \infty$ limit, is thus far an unproven conjecture.

These additional properties are, in the first place, the finiteness, in the $N \to \infty$ limit, of the coefficients $E_n(N, \rho)$ of Eq. (6) divided by N:

$$\lim_{N \to \infty} [E_n(N, \rho)/N] = \tilde{\epsilon}_n(\rho). \tag{8}$$

This property is a precondition for raising the issue of the convergence of the expansion

$$\sum_{n=0}^{\infty} \tilde{\epsilon}_n(\rho) g^n = \tilde{\epsilon}(\rho, g). \tag{9}$$

As we shall presently show, it may happen that this series converges; and the function $\tilde{\epsilon}(\rho, g)$ it defines may (but need not) coincide with the function $\tilde{\epsilon}(\rho, g)$ defined above.

The second additional property of the coefficients $E_n(N, \rho)$ is validity of the bound[8]

$$|E_n(N, \rho)| < \Gamma N n! (\tilde{\Lambda})^n \tag{10}$$

with Γ and $\tilde{\Lambda}$ independent of N.

Consider now the example

$$E(N; \rho, g) = -(1+z)Ng^2\{g^2 - (\rho - g)^2/$$
$$(1 + \exp[-N + N^2(g - \rho)])\} + zNg^2(g^2 - (\rho - g)^2/$$
$$\{1 + \exp[-N + N^2(g - \rho)]\})/[1 + \exp(N - gN^2)]$$
$$- g^2N^2/[1 + \exp(N + gN^2)]. \tag{11}$$

It is easily seen that it satisfies the inequality (10) and the condition (7), that it yields through (5)

$$\epsilon(\rho, g) = g^2\rho(\rho - 2g) \quad \text{for } \rho \geq \bar{\rho}(g) = g, \tag{12a}$$

$$\epsilon(\rho, g) = -g^4 = \bar{\epsilon}(g) \quad \text{for } \rho \leq \bar{\rho}(g) = g, \tag{12b}$$

and that the coefficients $\tilde{\epsilon}_n(\rho)$ of Eq. (8) are finite and yield, through Eq. (9),

$$\tilde{\epsilon}(\rho, g) = (1 + z)g^2\rho(\rho - 2g). \tag{13}$$

Note that the function $\tilde{\epsilon}(\rho, g)$ is entire and coincides, if $z = 0$, with the function $\epsilon(\rho, g)$ for $\rho \geq \bar{\rho}(g) = g > 0$ [so that $\tilde{\epsilon}(\rho, g) = \bar{\epsilon}(\rho, g)$]. Indeed, we have constructed our example so that, in this case ($z = 0$), $\bar{\epsilon}(g) = -g^4$ and $\bar{\rho}(g) = g$ could be recovered from $\tilde{\epsilon}(\rho, g)$ by applying the usual prescription, i.e., minimizing with respect to ρ.

Although we have built into the expression (11) of $E(N; \rho, g)$ some properties that are presumably also valid in realistic cases, it should be emphasized that the only purpose of this counterexample is to invalidate the claim that, using only the information discussed above, it is possible to prove that the energy per particle of the N-body system in the $N \to \infty$ limit is nonanalytic in the coupling constant g. The question of plausibility—whether it is likely that in realistic cases one or

the other possibility prevail—is outside of the scope of the present paper.[9] It is, however, worth calling attention to the mechanism whereby, in the example given here, singularities of the N-body system ground-state energy $E(N; \rho, g)$ approach, as N increases, the origin of the complex g plane, and indeed accumulate there [and also on the positive real axis, in the complex ρ plane, at the point corresponding to the actual density $\bar{\rho}(g)$ of the macroscopic system], but disappear in the $N \to \infty$ limit.[10] In the example displayed these singularities are of polar type[10]; it is clear that analogous examples with branch points instead of poles could be invented just as easily.

ACKNOWLEDGMENT

One of us (FC) would like to thank Professor G.A. Baker, Jr. for extensive correspondence on the subject of this paper.

[1]For conditions on the potential function $v(r)$ that are sufficient to guarantee saturation, see: D. Ruelle, *Statistical Mechanics—Rigorous Results* (Benjamin, New York, 1969); F. Calogero, Yu.A. Simonov, and E.L. Surkov, Phys. Rev. 5, 1493 (1972). Conditions that are necessary for saturation are also given in Ruelle's book; see also the review paper by F. Calogero and Yu.A. Simonov in *The Nuclear Many-Body Problem* (F. Calogero and C. Ciofi degli Atti, eds.) (Editrice Compositori, Bologna, 1974), Vol. I, p. 51; and the literature quoted there.

[2]We are therefore assuming the potential $v(r)$ to be repulsive at short range but attractive at longer range, as it is actually the case in most problems of physical interest.

[3]There might be breaks in the rising part of the curve, connected with phase transitions between different spatial arrangements; but these phenomena are outside of the scope of our discussion.

[4]The standard definition of a regular potential $v(r)$ requires that it be finite-valued for $r > 0$, less singular than r^{-2} at the origin ($\lim_{r \to 0} r^{2-\epsilon} v(r) = 0$, $\epsilon > 0$) and asymptotically integrable ($\lim_{r \to \infty} r^{3+\epsilon} v(r) = 0$, $\epsilon > 0$). Hereafter we limit our consideration to such potentials.

[5]In principle, of course, this is no limitation, since in any case we are primarily interested in the function $\bar{\epsilon}(g)$ for $g \approx g_0$, where g_0 is the physical value of the coupling constant; and eventually in the analytic continuation of this function to other values of g, including $g \approx 0$, even if for such values of g it looses any physical meaning.

[6]Note that this occurs, in three-dimensional space, even if the particles are identical fermions; see Ref. 1.

[7]G.A. Baker, Jr., Rev. Mod. Phys. 43, 479 (1971).

[8]Actually, the result that is proved in Ref. 7 is considerably less stringent, corresponding to Eq. (10) but with $(2n)!$ in place of $n!$ [see Eq. (3.38), and the sentence following it, in Ref. 7]. The conjecture that the strongest condition (10) hold is, however, made plausible [see Eq. (3.45), and the discussion leading to it and following it, in Ref. 7].

[9]Let us, however, mention that an example that is less artificial, being related to a (one-dimensional) mathematical model that displays some of the typical features of an N-body system, does reproduce the basic properties of the example given here: F. Calogero and A. Degasperis, J. Math. Phys. (to be published).

[10]The function $E(N; \rho, g)$ of Eq. (11) is clearly meromorphic in ρ and g; its poles in the g plane occur at $g = \pm g_n$ and at $g = \rho + g_n$, and those in the ρ plane occur at $\rho = g - g_n$, with $g_n = N^{-1} + i\pi(2n+1)N^{-2}$, $n = 0, \pm 1, \pm 2, \cdots$.

A possible unification of gravitation and electrodynamics both for spinless and spinning media

M. N. Mahanta

Department of Mathematics, Indian Institute of Technology, New Delhi-110029, India
(Received 29 August 1974)

It was shown by the author recently that the Einstein–Mayer formalism of 1931–32 can be successfully applied to derive the field equations for a spinning medium. The present article discusses the possibility of obtaining a unified representation of gravitation as well as electromagnetism both for spinless and spinning media with the help of the same formalism. Field equations and equations of motion for a point particle are derived for both cases.

1. INTRODUCTION

In 1931 Einstein and Mayer[1] published a unified theory of gravitation and electromagnetism based on a five-dimensional formalism. This theory yields the Einstein–Maxwell field equations correctly but only for the vacuum case. In a subsequent paper[2] in 1932 the authors sought to remedy this shortcoming by further generalizing the mathematical formalism. But as it has been recently shown by the author of this article,[3,4] the latter generalization correctly represents the intrinsic angular momentum of a spinning medium and has nothing to do with corpuscles as proposed by Einstein and Mayer. In this paper we shall first try to show that we have only to introduce a slight modification in Ref. 1 in order to get the desired representation of a nonzero current density vector and so derive the Einstein–Maxwell system of field equations for the nonempty case from a single formalism. The equations of motion for a charged particle are derived and they are essentially the Lorentz equations.

In the second part we introduce the spin of the medium as in Refs. 3, 4 and derive the combined system of field equations in the presence of gravitation, electrodynamics and spin and these equations in turn are used to derive the equations of motion of a charged spinning particle, which bear formal resemblance to the same equations in already existing literature.[5–9]

2. CASE OF A SPINLESS CHARGED MEDIUM

A. Mathematical formalism

The mathematical formalism is the same as in Ref. 4, except that we now assume as in Ref. 1, that

$$\gamma^{\rho}_{k;q} = A^{\sigma} F_{kq} \tag{2.1}$$

where F_{kq} is defined this way. We assume F_{kq} to be skew-symmetric. From this and the relation

$$A^{\sigma}\gamma^{k}_{\sigma} = 0, \tag{2.2}$$

it follows that

$$A_{\sigma;q} = -\gamma^{k}_{\sigma} F_{kq}. \tag{2.3}$$

It is also easily verified that

$$g_{lk;q} \equiv (\gamma_{\sigma l}\gamma^{\sigma}_{k})_{;q} = 0, \tag{2.4}$$

$$g^{\rho\sigma}_{;q} \equiv (A^{\rho}A^{\sigma} + \gamma^{\rho}_{k}\gamma^{\sigma k})_{;q} = 0. \tag{2.5}$$

Following the same procedure as in Ref. 4 we get the following expression for $P_{q\rho\lambda}{}^{\tau}$, viz.,

$$P_{q\rho\lambda}{}^{\tau} = -\gamma_{\lambda r}A^{\tau}(F_{\rho}{}^{r}{}_{;q} - F_{q}{}^{r}{}_{;\rho}) + \gamma^{\tau k}A_{\lambda}(F_{kq;\rho} - F_{k\rho;q}) + \gamma_{\lambda r}\gamma^{\tau k}(R_{k\rho q}{}^{r} + F_{kq}F_{\rho}{}^{r} - F_{k\rho}F_{q}{}^{r}), \tag{2.6}$$

from which

$$P_{\sigma\rho} \equiv \gamma^{q}_{\tau}P_{q\rho\sigma}{}^{\tau} = A_{\sigma}F_{\rho}{}^{k}{}_{;k} + \gamma_{\sigma r}(R^{r}_{\rho} - F_{k\rho}F^{kr}), \tag{2.7}$$

$$P \equiv \gamma^{\sigma\rho}P_{\sigma\rho} = R - F_{k\rho}F^{k\rho}. \tag{2.8}$$

We now form the tensor

$$U_{\sigma\rho} \equiv P_{\sigma\rho} - \tfrac{1}{4}\gamma_{\sigma\rho}(P + R) = A_{\sigma}F_{\rho}{}^{k}{}_{;k} + \gamma^{r}_{\sigma}[(R_{r\rho} - \tfrac{1}{2}g_{r\rho}R) - (F_{kr}F^{k}{}_{\rho} - \tfrac{1}{4}g_{r\rho}F_{kl}F^{kl})]. \tag{2.9}$$

It is then shown that

$$U^{\rho}_{\sigma;\rho} - \tfrac{1}{2}\gamma^{r}_{\sigma}N_{rk\rho}F^{k\rho} = 0 \tag{2.10}$$

where

$$N_{rk\rho} \equiv F_{rk,\rho} + F_{k\rho,r} + F_{\rho r,k}. \tag{2.11}$$

Thus

$$U^{\rho}_{\sigma;\rho} = 0 \quad \text{if } N_{rk\rho} = 0. \tag{2.12}$$

B. Field equations

We propose the following set of equations as field equations for the combined Einstein–Maxwell field, viz.,

$$U^{\sigma\rho} = KT^{\sigma\rho}, \tag{2.13}$$

$$N_{rk\rho} \equiv F_{rk,\rho} + F_{k\rho,r} + F_{\rho r,k} = 0. \tag{2.14}$$

$T^{\sigma\rho}$ is the "mixed" energy–momentum tensor and K is a universal constant. For a charged dust

$$T^{\sigma\rho} = \mu u^{\sigma}u^{\rho}, \tag{2.15}$$

μ being the mass density, u^{ρ} the four-dimensional velocity vector with u^{σ} as the corresponding 5-vector, i.e.,

$$u^{\sigma} = \gamma^{\sigma}_{k}u^{k} + (u_{\tau}A^{\tau})A^{\sigma}. \tag{2.16}$$

Writing (2.13) fully we get

$$A^{\sigma}F^{\rho k}{}_{;k}+\gamma^{\sigma r}[(R^{\rho}_{r}-\tfrac{1}{2}\delta^{\rho}_{r}R)-(F_{kr}F^{k\rho}-\tfrac{1}{4}\delta^{\rho}_{r}F_{kl}F^{kl})]$$
$$=K\mu\gamma^{\sigma}_{k}u^{k}u^{\rho}+K\mu\Lambda A^{\sigma}u^{\rho} \tag{2.17}$$

writing

$$\Lambda=A_{\tau}u^{\tau}. \tag{2.18}$$

Multiplying (2.17), respectively, by $\gamma_{\sigma q}$ and A_{σ}, we get

$$(R^{q\rho}-\tfrac{1}{2}g^{q\rho}R)-(F_{k}{}^{q}F^{k\rho}-\tfrac{1}{4}g^{q\rho}F_{kl}F^{kl})=K\mu u^{\rho}u^{q}, \tag{2.19}$$

$$F^{\rho k}{}_{;k}=(K\Lambda\mu)u^{\rho}. \tag{2.20}$$

If now we put

$$F_{ik}=\sqrt{K}\,\phi_{ik} \tag{2.21}$$

where ϕ_{ik} is the tensor of the electromagnetic field (K is positive from the Einstein theory with our definition of $R^{q\rho}$), then we have the following system of Einstein–Maxwell equations:

$$R^{q\rho}-\tfrac{1}{2}g^{q\rho}R=K[\mu u^{q}u^{\rho}+(\phi_{kq}\phi^{k\rho}-\tfrac{1}{4}g^{q\rho}\phi_{kl}\phi^{kl})], \tag{2.22}$$

$$\phi^{\rho k}{}_{;k}=\rho u^{\rho}, \tag{2.23}$$

where $\rho\equiv\sqrt{K}\Lambda\mu$ is defined as the charge density:

$$\phi_{rk,\rho}+\phi_{k\rho,r}+\phi_{\rho r,k}=0. \tag{2.24}$$

C. The Lorentz equations of motion for a charged test particle

The Lorentz equations for a charged particle can be obtained by taking the divergence of (2.22) and using (2.23) and putting $\rho/\mu=e/m$, e and m being the charge and mass of the particle, respectively. We may derive the same equation directly from (2.13) since

$$(T^{\sigma\rho})_{;\rho}=\frac{1}{K}U^{\sigma\rho}{}_{;\rho}\equiv 0,$$

i.e.,

$$(\mu u^{\rho}u^{\sigma})_{;\rho}=0, \tag{2.25}$$

i.e.,

$$(\mu u^{\rho})_{;\rho}u^{\sigma}+\mu u^{\rho}u^{\sigma}{}_{;\rho}=0, \tag{2.26}$$

which yields $(\mu u^{\rho})_{;\rho}=0$ if $u_{\sigma}u^{\sigma}=\text{const}$. Thus from (2.26) we get

$$u^{\sigma}_{;\rho}u^{\rho}=0,$$

i.e.,

$$\frac{\mathfrak{D}u^{\sigma}}{\mathfrak{D}\tau}=0. \tag{2.27}$$

From $u^{k}=\gamma^{k}_{\sigma}u^{\sigma}$, we thus get

$$\frac{\mathfrak{D}u^{k}}{\mathfrak{D}\tau}=\gamma^{k}_{\sigma;r}u^{\sigma}u^{r},$$

i.e.,

$$\frac{du^{k}}{d\tau}+\begin{Bmatrix}k\\ p\ q\end{Bmatrix}u^{\rho}u^{q}=A_{\sigma}u^{\sigma}F^{k}{}_{r}u^{r}$$
$$=\Lambda\sqrt{K}\phi^{k}{}_{r}u^{r}$$
$$=\frac{\rho}{\mu}\phi^{k}{}_{r}u^{r} \tag{2.28}$$

which is the covariant form of Lorentz equations replacing ρ/μ by e/m.

We may easily prove that $\Lambda=A_{\sigma}u^{\sigma}$ is conserved along the world line of the charge, since

$$\frac{\mathfrak{D}(\Lambda)}{\mathfrak{D}\tau}\equiv\frac{\mathfrak{D}(A_{\sigma}u^{\sigma})}{\mathfrak{D}\tau}=A_{\sigma;\rho}u^{\rho}u^{\sigma}+A_{\sigma}\frac{\mathfrak{D}u^{\sigma}}{\mathfrak{D}\tau}$$
$$=\gamma^{k}_{\sigma}F_{\rho k}u^{\rho}u^{\sigma}=F_{\rho k}u^{\rho}u^{k}=0.$$

3. CASE OF A CHARGED SPINNING MEDIUM

A. Mathematical formalism

To take account of the internal angular momentum (spin) of the medium we further generalize (2.1) by taking

$$\gamma^{\sigma}_{k;q}=A^{\sigma}F_{kq}+\gamma^{\sigma r}V_{rkq}. \tag{3.1}$$

It is then easily obtained that

$$A^{\sigma}_{;q}=-\gamma^{\sigma k}F_{kq}. \tag{3.2}$$

For

$$g_{lk;q}\equiv(\gamma_{\sigma l}\gamma^{\sigma}_{k})_{;q}=0$$

we must have

$$V_{rkq}=-V_{krq} \tag{3.3}$$

and then

$$g^{\rho\sigma}_{;q}\equiv(A^{\rho}A^{\sigma}+\gamma^{\rho}_{k}\gamma^{\sigma k})_{;q}=0.$$

It is to be noted that the relation $A^{\sigma}_{;q}=0$ used in Ref. 4 follows from (3.2) if $F_{kq}=0$.

We are led to the following expression for $P_{q\rho\sigma}{}^{\tau}$, viz.,

$$P_{q\rho\sigma}{}^{\tau}=\gamma_{\sigma r}A^{\tau}(F^{r}{}_{\rho;q}-F^{r}{}_{q;\rho}+V^{r}{}_{kq}F^{k}{}_{\rho}-V^{r}{}_{k\rho}F^{k}{}_{q})$$
$$+\gamma^{\tau k}A_{\sigma}(F_{kq;\rho}-F_{k\rho;q}+F^{r}{}_{\rho}V_{rkq}-F^{r}{}_{q}V_{rk\rho})$$
$$+\gamma_{\sigma r}\gamma^{\tau k}(R_{k\rho q}{}^{r}-F_{kq}F^{r}{}_{\rho}+F^{r}{}_{q}F_{k\rho}+V^{rs}{}_{\rho}V_{skq}$$
$$-V^{rs}{}_{q}V_{sk\rho}+V^{r}{}_{kq;\rho}-V^{r}{}_{k\rho;q}). \tag{3.4}$$

Then

$$P_{\sigma\rho}\equiv\gamma^{q}_{\tau}P_{q\rho\sigma}{}^{\tau}=A_{\sigma}(-F^{q}{}_{\rho;q}+F^{r}{}_{\rho}V_{r}{}^{q}{}_{q}-F^{rk}V_{rk\rho})$$
$$+\gamma_{\sigma r}(R^{r}_{\rho}-F^{kr}F_{k\rho}+V^{rs}{}_{\rho}V_{s}{}^{q}{}_{q}-V^{rsk}V_{sk\rho}$$
$$+V^{rq}{}_{q;\rho}-V^{rq}{}_{\rho;q}), \tag{3.5}$$

$$P\equiv\gamma^{\sigma\rho}P_{\sigma\rho}=[R-F^{k\rho}F_{k\rho}+V^{\rho r}{}_{\rho}V_{r}{}^{q}{}_{q}-V^{\rho sk}V_{sk\rho}$$
$$+V^{\rho q}{}_{q;\rho}-V^{\rho q}{}_{\rho;q}], \tag{3.6}$$

$$U_{\sigma\rho}\equiv P_{\sigma\rho}-\tfrac{1}{4}\gamma_{\sigma\rho}(P+R)$$
$$=A_{\sigma}[F_{\rho}{}^{k}{}_{;k}+F_{r\rho}V^{rq}{}_{q}-F_{rk}V^{rk}{}_{\rho}]$$
$$+\gamma^{r}_{\sigma}[(R_{r\rho}-\tfrac{1}{2}g_{r\rho}R)-(F_{kr}F^{k}{}_{\rho}-\tfrac{1}{4}g_{r\rho}F_{kl}F^{kl})$$
$$+(V_{rs\rho}V^{sq}{}_{q}-V_{r}{}^{sk}V_{sk\rho}+V_{r}{}^{q}{}_{q;\rho}-V_{r}{}^{q}{}_{\rho;q})$$
$$-\tfrac{1}{4}g_{r\rho}(V^{lr}{}_{l}V_{r}{}^{q}{}_{q}-V^{lsk}V_{skl}+V^{lq}{}_{q;l}-V^{lq}{}_{l;q})]. \tag{3.7}$$

B. Field equations

We shall consider an incoherent charged distribution of spinning matter and restrict ourselves to the case when the electric dipole moment density is zero and the magnetic dipole moment density m^{ij} is parallel to the spin density s^{ij}, i.e.,

$$m^{ij}=\lambda s^{ij} \tag{3.8}$$

and

$$m^{ij}u_j = \lambda s^{ij}u_j = 0. \tag{3.9}$$

The "mixed" energy–momentum tensor will have the form

$$T^{\sigma\rho} = u^\sigma g^\rho + m^{\sigma k}\phi_k{}^\rho \tag{3.10}$$

where g^ρ is the momentum density.

The additional term on the rhs is the interaction energy between the dipole moment density and the electromagnetic field represented by the skew-symmetric tensor $\phi_{k\rho}$. Thus our field equations are

$$U^{\sigma\rho} = KT^{\sigma\rho} = K[u^\sigma g^\rho - m^{\sigma k}\phi^\rho{}_k]. \tag{3.11}$$

Writing

$$u^\sigma = \gamma^\sigma_q u^q + (A_\lambda u^\lambda)A^\sigma$$

and

$$m^{\sigma k} = \gamma^\sigma_q m^{qk} + (A_\lambda m^{\lambda k})A^\sigma$$

in (3.11) and substituting for $U^{\sigma\rho}$ and multiplying both sides by A_σ and γ^r_σ, respectively, we get

$$F^{\rho k}{}_{;k} - F_{rk}V^{rk\rho} = K[(A_\sigma u^\sigma)g^\rho - (A_\lambda m^{\lambda k})\phi^\rho{}_k], \tag{3.12}$$

$$(R^{\rho q} - \tfrac{1}{2}g^{\rho q}R) - V^{qsk}V_{sk}{}^\rho - V^{ql\rho}{}_{;l} + \tfrac{1}{4}g^{\rho q}V^{lsk}V_{skl}$$

$$= K(g^\rho u^q - \phi^\rho{}_k m^{qk}) + (F^{\rho k}F^q{}_k - \tfrac{1}{4}g^{\rho q}F_{kl}F^{kl}). \tag{3.13}$$

In deriving (3.12) and (3.13) we have already assumed the following relations:

$$V^{lij} = -\tfrac{1}{2}K(u^i s^{lj} + u^j s^{li} - u^l s^{ij}) \tag{3.14}$$

and

$$V^{li}{}_i = V^{li}{}_l = 0 \quad \text{since } s^{ij}u_j = 0. \tag{3.15}$$

If we now set

$$F_{ik} = \sqrt{K}\phi_{ik},$$

then (3.12) reduces to

$$\phi^{\rho k}{}_{;k} = \sqrt{K}(A_\sigma u^\sigma)g^\rho + \phi_{rk}V^{rk\rho} - (A_\lambda m^{\lambda k})F^\rho{}_k.$$

Now

$$(A_\lambda m^{\lambda k})F^\rho{}_k = m^{\lambda k}(\gamma^\rho_{\lambda;k} - \gamma^r_\lambda V_r{}^\rho{}_k)$$

$$= (m^{\lambda k}\gamma^\rho_\lambda)_{;k} - \gamma^\rho_\lambda(m^{\lambda k})_{;k} - m^{\lambda k}\gamma^r_\lambda V_r{}^\rho{}_k$$

$$= m^{\rho k}{}_{;k} - \gamma^\rho_\lambda(m^{\lambda k})_{;k} + m_{rk}V^{\rho rk}$$

Considering, as in Ref. 4, the middle term as the covariant divergence of $m^{\rho k}$ w.r.t a new connection Γ and using Eq. (3.12) of Ref. 4, we get the rhs as zero. Thus we get

$$\phi^{\rho k}{}_{;k} = j^\rho \tag{3.16}$$

with $j^\rho = \sqrt{K}(A_\sigma u^\sigma)g^\rho + \phi_{rk}V^{rk\rho}$ as a generalized current density vector. Equation (3.16) is the analog of Maxwell's equations in our case. If we put $s^{ij} = 0$, then we get back the familiar form of Maxwell's equations.

Equation (3.13) now reads

$$(R^{\rho q} - \tfrac{1}{2}g^{\rho q}R) - V^{qsk}V_{sk}{}^\rho + V^{lq\rho}{}_{;l} + \tfrac{1}{4}g^{\rho q}V^{lsk}V_{skl}$$

$$= K[(g^\rho u^q - \phi^\rho{}_k m^{qk}) + (\phi^{\rho k}\phi^q{}_k - \tfrac{1}{4}g^{\rho q}\phi_{kl}\phi^{kl})], \tag{3.17}$$

which are analogs of Einstein's field equations. Taking the skew-symmetric part of (3.17), we get

$$(V^{lq\rho} - V^{l\rho q})_{;l} = K[g^\rho u^q - g^q u^\rho - (\phi^\rho{}_k m^{qk} - \phi^q{}_k m^{\rho k})],$$

i.e.,

$$(s^{\rho q}u^l)_{;l} = [g^\rho u^q - g^q u^\rho - (\phi^\rho{}_k m^{qk} - \phi^\rho{}_k m^{qk})]. \tag{3.18}$$

C. Equations of motion of a charged spinning point particle

Using the familiar procedure of integrating over the small volume V_0 of the particle, we get from Eq. (3.18) first,

$$\dot{S}^{\rho q} = G^\rho u^q - G^q u^\rho - (\phi^\rho{}_k M^{qk} - \phi^q{}_k M^{\rho k}) \tag{3.19}$$

where the capital letters are the integrated small letters over V_0 and the dot is differentiation w.r.t proper time. To get the other set of equations, we first take the divergence of Eq. (3.17) ignoring the terms quadratic in spin and then integrate over V_0. We shall then get

$$\dot{G}^\rho = -\tfrac{1}{2}u^l S^{qr}R_{lqr}{}^\rho + \phi^\rho{}_k J^k - \phi^\rho{}_{k;q}M^{kq} - \phi^\rho{}_k M^{kq}{}_{;q}. \tag{3.20}$$

The only essentially new term on the rhs is the last term.

[1] A. Einstein and W. Mayer, Sitzungsber Preuss Akad. Wiss., Phys. Math. Kl. 541 (1931).
[2] A. Einstein and W. Mayer, Sitzungsber Preuss Akad. Wiss., Phys. Math. Kl., 130 (1932).
[3] M.N. Mahanta "On a possible generalization of the field equations of general relativity for the case of a spinning medium," Seminar at the Institute Henri Poincaré, Paris, 1972.
[4] M.N. Mahanta and A. Sudhakar Rao, J. Math. Phys. 16, 381 (1975).
[5] W. Israel, Lett. Nuovo Cimento 7, 16 (1973).
[6] W.G. Dixon, Proc. Roy. Soc. A 319, 509 (1970).
[7] J. Madore, Ann. Inst. Henri Poincaré 11, 221 (1969).
[8] L.G. Suttorp and S.R. De Groot, Nuovo Cimento, 65 A, 245 (1970).
[9] H.P. Künzle, Commun. Math. Phys. 27, 23 (1972).

Solutions of a scalar-tensor theory from Einstein's theory

T. Singh

Department of Mathematics, University of Gorakhpur, Gorakhpur 273001, India
(Received 30 December 1974)

The vacuum field equations of a scalar-tensor theory proposed by Sen and Dunn are investigated. The metric coefficients are assumed to be functions of three variables only. It is then shown that if one assumes a functional relationship between some one of the metric coefficients and the scalar function, one can find a solution of the vacuum field equations of Sen and Dunn's scalar-tensor theory in terms of a solution of the vacuum field equations of Einstein's theory. This result is applied to some well-known solutions, viz., Schwarzschild solution, the static plane symmetric solution of Taub, the conformastat solution of Das, and the static axially symmetric solutions of Levi-Civita and Curzon; and the corresponding solutions of the scalar-tensor theory are obtained.

1. INTRODUCTION

Recently Sen and Dunn[1] have proposed a new scalar–tensor theory of gravitation in which both the scalar and the tensor fields have intrinsic geometrical significance. The scalar field in this theory is characterized by the function $\Phi = \Phi(x^i)$ where x^i are coordinates in the four-dimensional Lyra manifold and the tensor field is identified with the metric tensor g_{ij} of the manifold. The field equations given by Sen and Dunn for the combined scalar tensor fields are

$$R_{ij} - \tfrac{1}{2}g_{ij}R = \omega\Phi^{-2}(\Phi_{,i}\Phi_{,j} - \tfrac{1}{2}g_{ij}\Phi_{,k}\Phi^{,k}) - 8\pi G\Phi^{-2}T_{ij}, \tag{1.1}$$

where $\omega = 3/2$, T_{ij} is the energy–momentum tensor of the field, and R_{ij} and R are, respectively, the usual Ricci tensor and Riemann curvature scalar. [It may be noted that the field equations (1.1) do not follow from the original variational principle of Sen and Dunn. They can either be derived from a suitable action integral or assumed *a priori*.] In the matter-free region these field equations reduce to

$$R_{ij} - \tfrac{1}{2}g_{ij}R = \omega\Phi^{-2}(\Phi_{,i}\Phi_{,j} - \tfrac{1}{2}g_{ij}\Phi_{,k}\Phi^{,k}), \tag{1.2}$$

where $\omega = 3/2$.

In this paper we have investigated the vacuum field equations (1.2) of this scalar–tensor theory and discussed a procedure which will enable one to construct exact solutions of the scalar–tensor theory from given vacuum solutions of Einstein's theory. We assume that the tensor field g_{ij} and the scalar field Φ are functions of x^1, x^2, x^3 only. It is then shown that if one assumes a functional relationship between g_{00} and the scalar field Φ, one can find a solution of the scalar–tensor theory in terms of a solution of Einstein's vacuum field equations.

In Sec. 2 we first set up the field equations of the scalar–tensor theory in a suitable form by assuming a functional relationship between g_{00} and the scalar field Φ. Then we prove the desired result. In Sec. 3 we illustrate the application of the result for some well-known vacuum solutions of Einstein's theory, viz., Schwarzschild's exterior solution both in *standard* and in *isotropic* coordinates, conformastat solution of Das,[2] the static plane symmetric solution of Taub,[3] and the static axially symmetric solutions of Levi-Civita[4] and Curzon.[5] The last section contains some concluding remarks.

2. DERIVATION OF THE SCALAR TENSOR FIELDS FROM EINSTEIN VACUUM FIELD

We work with the vacuum field equations (1.2) of Sen and Dunn's scalar–tensor theory[1] which can be written in the form

$$R_{ij} = \omega\Phi^{-2}(\Phi_{,i}\Phi_{,j}), \tag{2.1}$$

where $\omega = 3/2$. This can be transformed into the equation

$$R_{ij} = \omega(h_{,i}h_{,j}), \tag{2.2}$$

where we have put

$$\Phi = e^h. \tag{2.3}$$

Further, we choose the matric in the form

$$ds^2 = e^{2U}(dx^0)^2 + e^{-2U}(\gamma_{\alpha\beta}dx^\alpha dx^\beta). \tag{2.4}$$

This metric is the most general static line element and admits a hypersurface orthogonal Killing vector $\xi^i = \delta^i_0$. The three-dimensional metric satisfies the following

$$\gamma_{\alpha\beta}\gamma^{\alpha\sigma} = \delta^\sigma_\beta. \tag{2.5}$$

The components of the Ricci tensor for the metric (2.4) are given by

$$R_{00} = e^{4U}\Delta_2(U), \tag{2.6}$$

$$R_{0\alpha} = 0, \tag{2.7}$$

$$R_{\alpha\beta} = P_{\alpha\beta} + 2U_{,\alpha}U_{,\beta} - \gamma_{\alpha\beta}\Delta_2(U), \tag{2.8}$$

where we have introduced the differential parameters of the first and second order[6] defined as

$$\Delta_1(U) = \gamma^{\alpha\beta}U_{,\alpha}U_{,\beta}, \tag{2.9}$$

$$\Delta_1(U, V) = \gamma^{\alpha\beta}U_{,\alpha}V_{,\beta}, \tag{2.10}$$

$$\Delta_2(U) = \gamma^{\alpha\beta}U_{;\alpha\beta} = \gamma^{\alpha\beta}\left(U_{,\alpha\beta} - \sum_{\alpha\beta}^{\sigma} U_{,\sigma}\right). \tag{2.11}$$

$\sum^\sigma_{\alpha\beta}$ and $P_{\alpha\beta}$ are respectively the Christoffel symbols and Ricci tensor defined with respect to the 3-space metric $\gamma_{\alpha\beta}$. The scalar field Φ and consequently h also will be taken as a function of x^1, x^2, x^3 only.

Now the field equations (1.2) for the metric (2.4) can easily be set up and one obtains

$$\Delta_2(U) = 0, \tag{2.12}$$

$$P_{\alpha\beta} + 2U_{,\alpha}U_{,\beta} = \omega(h_{,\alpha}h_{,\beta}). \tag{2.13}$$

We now assume that U and h are functionally related and the relation is of the form

$$U = bh, \tag{2.14}$$

where b is a constant.

If we assume (2.14) then the Eqs. (2.12) and (2.13) yield

$$\Delta_2(h) = \gamma^{\alpha\beta}h_{;\alpha\beta} = 0 \tag{2.15}$$

and

$$P_{\alpha\beta} + 2k^2 h_{,\alpha}h_{,\beta} = 0, \tag{2.16}$$

where $k^2 = b^2 - \omega/2$. Now if we write

$$f = kh, \tag{2.17}$$

then Eqs. (2.15) and (2.16) reduce to

$$P_{\alpha\beta} + 2f_{,\alpha}f_{,\beta} = 0, \tag{2.18}$$

$$\Delta_2(f) = \gamma^{\alpha\beta}f_{;\alpha\beta} = 0. \tag{2.19}$$

Equations (2.18) and (2.19) are nothing but the static gravitational field equations $R_{ij} = 0$ of Einstein's theory in empty space; where the line element has been taken to be

$$ds^2 = e^{2f}(dx^0)^2 + e^{-2f}(\gamma_{\alpha\beta}dx^\alpha dx^\beta). \tag{2.20}$$

Thus we have established the following result:

> For every solution f and $\gamma_{\alpha\beta}$ of the empty space field equations of Einstein's theory, we can construct a corresponding solution of the vacuum field equations of Sen and Dunn's scalar–tensor theory with the same $\gamma_{\alpha\beta}$, Φ being given by (2.3) and (2.17) and U being given by (2.14) and (2.17).

In the derivation of the above result we have assumed that all the functions are independent of the coordinate x^0. However, the result is quite general and holds good with appropriate modifications when the absent coordinate may be any of the four variables x^0, x^1, x^2, x^3.

3. APPLICATIONS OF THE TECHNIQUE

Now applying the technique developed in the previous section, we shall obtain the solutions of the vacuum field equations of Sen and Dunn's scalar–tensor theory from the known solutions of vacuum field equations $R_{ij} = 0$ of Einstein's theory of gravitation.

A. Schwarzschild's solution in standard coordinates

This is given by the line element

$$\begin{aligned} ds^2 = {} & \exp[\log(1 - 2m/r)]\,dt \\ & - \exp[-\log(1 - 2m/r)][dr^2 + r^2(1 - 2m/r) \\ & \times (d\theta^2 + \sin^2\theta\, d\phi^2)]. \end{aligned} \tag{3.1}$$

The corresponding solution of the scalar–tensor theory is

$$\begin{aligned} ds^2 = {} & \left(1 - \frac{2m}{r}\right)^{b/\sqrt{b^2-\omega/2}} dt^2 \\ & - \left(1 - \frac{2m}{r}\right)^{-b/\sqrt{b^2-\omega/2}} \\ & \times \left[dr^2 + r^2\left(1 - \frac{2m}{r}\right)(d\theta^2 + \sin^2\theta\, d\phi^2)\right] \end{aligned} \tag{3.2}$$

with the scalar function Φ given by

$$\Phi = \left(1 - \frac{2m}{r}\right)^{1/2\sqrt{b^2-\omega/2}}. \tag{3.3}$$

B. Schwarzschild's solution in isotropic coordinates

This is given by the metric

$$\begin{aligned} ds^2 = {} & \left(\frac{1 - m/r}{1 + m/r}\right)^2 dt^2 - \left(\frac{1 - m/r}{1 + m/r}\right)^{-2} \\ & \times \left(1 - \frac{m^2}{r^2}\right)^2 (dr^2 + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\phi^2). \end{aligned} \tag{3.4}$$

The corresponding solution of the scalar–tensor theory is

$$\begin{aligned} ds^2 = {} & \left(\frac{1 - m/r}{1 + m/r}\right)^{2b/\sqrt{b^2-\omega/2}} dt^2 - \left(\frac{1 - m/r}{1 + m/r}\right)^{-2b/\sqrt{b^2-\omega/2}} \\ & \times \left[\left(1 - \frac{m^2}{r^2}\right)(dr^2 + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\phi^2)\right] \end{aligned} \tag{3.5}$$

with the scalar function

$$\Phi = \left(\frac{1 - mr}{1 + mr}\right)^{1/\sqrt{b^2-\omega/2}}. \tag{3.6}$$

If we write $m = B$, $b = -1/C$ and $\sqrt{b^2 - \omega/2} = -\lambda/C$, then (3.5) and (3.6) transform, respectively, to

$$\begin{aligned} ds^2 = {} & \left(\frac{1 - B/r}{1 + B/r}\right)^{2/\lambda} dt^2 - \left[\left(1 - \frac{B}{r}\right)^{1-1/\lambda}\left(1 + \frac{B}{r}\right)^{1+1/\lambda}\right] \\ & \times (dr^2 + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\phi^2) \end{aligned} \tag{3.7}$$

and

$$\Phi = \left(\frac{1 - B/r}{1 + B/r}\right)^{-C/\lambda}, \tag{3.8}$$

where B, C, and λ are constants. This solution is a particular case of the spherically symmetric solution in isotropic coordinates obtained by Halford[7] when we choose $p_0 = q_0 = 0$, $x_0^0 = 1$ [see Eqs. (2.42a), (2.42b), (2.42c), and (2.43) of Ref. 7].

C. A conformastat solution

Das[2] has obtained a conformastat solution

$$ds^2 = (1 - mx)^{-2}\, dt^2 - (1 - mx)^4(dx^2 + dy^2 + dz^2), \tag{3.9}$$

where m = const. This metric is due to an infinite plate parallel to the (y, z)-plane. The corresponding solution of the scalar–tensor theory is

$$\begin{aligned} ds^2 = {} & (1 - mx)^{-2b/\sqrt{b^2-\omega/2}}\, dt^2 \\ & - (1 - mx)^{2+(2b/\sqrt{b^2-\omega/2})}(dx^2 + dy^2 + dz^2) \end{aligned} \tag{3.10}$$

with the scalar function

$$\Phi = (1 - mx)^{-1/\sqrt{b^2-\omega/2}}. \tag{3.11}$$

D. A static plane symmetric solution

The static plane symmetric solution of Taub[3] is

$$ds^2 = (k_1x + k_2)^{-1/2}(dt^2 - dx^2) - (k_1x + k_2)(dy^2 + dz^2), \tag{3.12}$$

where k_1 and k_2 are constants. This is due to the gravitational field of an infinite plane parallel to the (y,z)-plane. The solution of the scalar—tensor theory is given by the metric

$$ds^2 = (k_1x + k_2)^{-b/2\sqrt{b^2-\omega}/2}\,dt^2 - (k_1x + k_2)^{b/2\sqrt{b^2-\omega}/2} \times [(k_1x + k_2)^{-1}\,dx^2 + (k_1x + k_2)^{1/2}(dy^2 + dz^2)] \tag{3.13}$$

together with the scalar function

$$\Phi = (k_1x + k_2)^{-1/4\sqrt{b^2-\omega}/2} \tag{3.14}$$

E. Levi-Civita solution

The static axially symmetric solution of Levi-Civita[4] is given by the metric

$$ds^2 = \left(\frac{r}{r_0}\right)^{(q^2+2q)/2}(dt^2 - dr^2) - \left(\frac{r}{r_0}\right)^{q} r^2\,d\phi^2 - \left(\frac{r}{r_0}\right)^{-q} dz^2, \tag{3.15}$$

where r_0 and q are constants. This is due to the gravitational field of a line mass placed along the z axis.

The use of the transformation $z \to it$, $t \to iz$ brings the metric to the standard form and then applying the result of the previous section the corresponding solution of the scalar—tensor theory is

$$ds^2 = \left(\frac{r}{r_0}\right)^{q^2/2+(bq/2\sqrt{b^2-\omega}/2)}(dt^2 - dr^2) - \left(\frac{r}{r_0}\right)^{bq/\sqrt{b^2-\omega}/2} r^2\,d\phi^2 - \left(\frac{r}{r_0}\right)^{-bq/\sqrt{b^2-\omega}/2} dz^2 \tag{3.16}$$

with the scalar function

$$\Phi = \left(\frac{r}{r_0}\right)^{-q/2\sqrt{b^2-\omega}/2} \tag{3.17}$$

F. "Curzon" particle solution

The static axially symmetric solution representing a "Curzon" particle is given by the metric

$$ds^2 = \exp\left(-\frac{2m}{\rho}\right)dt^2 - \exp\left(\frac{2m}{\rho}\right) \times \left[\exp\left(-\frac{m^2r^2}{2\rho^4}\right)(dr^2 + dz^2) + d\phi^2\right], \tag{3.18}$$

where $m = \text{const}$ and $\rho = (r^2 + z^2)^{1/2}$. The corresponding solution of the scalar—tensor theory is given by the metric

$$ds^2 = \exp(-2mb/\rho\sqrt{b^2 - \omega/2})\,dt^2 - \exp(2mb/\rho\sqrt{b^2 - \omega/2}) \times \left[\exp -\left(\frac{m^2r^2}{2\rho^4}\right)(dr^2 + dz^2) + d\phi^2\right] \tag{3.19}$$

together with the scalar function

$$\Phi = \exp(-m\rho/\sqrt{b^2 - \omega/2}). \tag{3.20}$$

4. CONCLUSION

The immediate use of the result obtained in Sec. 2 of this paper is in obtaining the solutions of Sen and Dunn's scalar tensor theory from the known empty spaces of Einstein's theory. But aside from the generation of solutions, one may also use the result obtained here in discussing the singularities of the two theories. One may easily observe that, $\gamma_{\alpha\beta}$ being the same in both the cases, the solutions in the two theories will have the same singularities, if any, present in $\gamma_{\alpha\beta}$.

In conclusion, we hope that the results of the present paper will lead to a deeper understanding of the relations between Einstein's theory of gravitation and the scalar—tensor theory of Sen and Dunn. We also hope that some physical insight can be gained from the solutions obtained in this paper.

ACKNOWLEDGMENT

The author is thankful to Professor K. B. Lal, Chairman of the Department of Mathematics, Gorakhpur University, for his keen interest in this work.

[1] D. K. Sen and K. A. Dunn, J. Math. Phys. **12**, 578 (1971).
[2] A. Das, J. Math. Phys. **12**, 1136 (1971).
[3] A. H. Taub, Ann. Math. **53**, 472 (1951).
[4] T. Levi-Civita, C. R. Acad. Lincei **28**, 101 (1919).
[5] H. E. J. Curzon, Proc. Lond. Math. Soc. **23**, 477 (1924).
[6] L. P. Eisenhart, *Riemannian Geometry* (Princeton U. P., Princeton, N. J., 1966), p. 41.
[7] W. D. Halford, J. Math. Phys. **13**, 1699 (1972).

Cores for operators in quantum mechanics

F. E. Schroeck Jr.*

Department of Mathematics, Florida Atlantic University, Boca Raton, Florida 33432
(Received 22 August 1974; revised manuscript received 11 February 1975)

It is shown that the set of finite linear combinations of the harmonic oscillator functions as well as the set of (Schwartz) functions of fast decrease form cores for a large class of operators on the space $L^2_\mu(\mathbb{R}^n)$, where μ is nice relative to Lebesgue measure. The results are applied to a variety of classes of classical potentials and their second-quantized analogs.

I. INTRODUCTION

In solving models of quantum field theory, one rapidly gets involved in tedious calculations in which the technical difficulties are partially alleviated if one can choose a nice core on which to calculate.[1] We present two such cores here, one consisting of finite linear combinations of the harmonic oscillator functions (well-known to most physicists) and the other consisting of the functions of fast decrease (well-known to most functional analysts).

II. THEORY

Definitions: (a) $\{\psi_j(x)\}$ denotes the set of harmonic oscillator (h.o.) functions in one variable; $\{\psi_\mathbf{j}(\mathbf{x})\}$, the h.o. functions in several variables obtained as decomposable tensor products of the one-variable h.o. functions:

$$\psi_\mathbf{j}(\mathbf{x}) = \psi_{j_1}(x_1)\psi_{j_2}(x_2)\cdots\psi_{j_n}(x_n), \quad \mathbf{j}=(j_1,\ldots,j_n).$$

D_0 denotes the set of finite linear combinations of h.o. functions.

(b) $\mathcal{S}(\mathbb{R}^n)$ denotes the set of functions of fast decrease.[2] The h.o. functions are in $\mathcal{S}(\mathbb{R}^n)$.[3]

(c) μ is to be a measure on $\mathbb{R}^n$, which is (i) absolutely continuous with respect to Lebesgue measure, (ii) such that the h.o. functions are in $L^2_\mu(\mathbb{R}^n)$, and (iii) such that for each compact subset S of $\mathbb{R}^n$, there exists a constant $d(S)$ such that

$$\|f\|_{L^2_\mu} \leq d(S)\|f\|_{L^2} \quad \text{for all } f\in L^2(\mathbb{R}^n),\ \operatorname{supp} f \subseteq S.$$

(d) A closable linear operator A on $L^2_\mu(\mathbb{R}^n)$ is said to have property P if there exist constants c, m such that

$$\|A\psi_\mathbf{j}\|_\mu \leq c(1+|\mathbf{j}|)^m, \quad |\mathbf{j}| = \sum_{i=1}^n j_i,$$

holds for all h.o. functions $\psi_\mathbf{j}$. The set of operators having property P is a linear set. In the case where μ = Lebesque measure, since the Fourier transform operator F sends h.o. functions to themselves and is unitary, A has property P iff FAF^{-1} has property P.

Theorem 1: Let A have property P, and let B denote any closed extension of A restricted to D_0. Then the domain of B contains $\mathcal{S}(\mathbb{R}^n)$.

The *Proof* follows immediately from the result[3] that any $f\in\mathcal{S}(\mathbb{R}^n)$ may be written in the form

$$f=\sum_\mathbf{j} a(\mathbf{j})\psi_\mathbf{j}, \quad a(\mathbf{j}) = \int \psi_\mathbf{j}(\mathbf{x}) f(\mathbf{x})\, d^n x,$$

and then $\Sigma_\mathbf{j}|a(\mathbf{j})|^2(1+|\mathbf{j}|)^q$ converges for all q.

More definitions: (e) Let α be a complex-valued Lebesgue-measurable function on $\mathbb{R}^n$ and define the closed set $\Gamma(\alpha)=\{\mathbf{x}\in\mathbb{R}^n \mid \alpha \text{ is essentially unbounded in all deleted neighborhoods of } \mathbf{x}\}$.

(f) $\{E_\mathbf{x}\}$ denotes the spectral family on $L^2(\mathbb{R})$ corresponding to the position operator. [By (d) above, this may be position in either configuration or momentum space.]

Theorem 2: Let A be an operator on $L^2_\mu(\mathbb{R}^n)$ which is multiplication by α ($A=\int \alpha(\mathbf{x})dE_\mathbf{x}$) with $\Gamma(\alpha)$ of Lebesgue measure zero (and hence μ-measure zero) and such that the domain of A contains $\mathcal{S}(\mathbb{R}^n)$. Then $\mathcal{S}(\mathbb{R}^n)$ is a core for A.

Proof: In fact we show that the functions of fast decrease with compact support form a core for A. To this end we introduce the closed sets $S(\delta)=\{\mathbf{x}\in\mathbb{R}^n \mid \text{distance } (\mathbf{x},\Gamma(\alpha)) \geq \delta,\ \|\mathbf{x}\|\leq\delta^{-1}\}$, for each $\delta>0$. Then since $\Gamma(\alpha)=\mathbb{R}^n\setminus\cup_{\delta>0}S(\delta)$ is of μ measure zero, for δ sufficiently small $S(\delta)$ is of nonzero μ-measure (and hence Lebesgue measure). α is essentially bounded on each $S(\delta)$, and so we introduce $D_1=\{f\in L^2_\mu(\mathbb{R}^n) \mid \operatorname{supp} f \subseteq S(\delta) \text{ for some } \delta>0\}$. D_1 is a stable dense set of analytic vectors for A and hence a core for A.[4] Now let B be any closed extension of A restricted to $\mathcal{S}(\mathbb{R}^n)$. It suffices to show that D_1 is in the domain of B. Let $f\in D_1$ be such that $\operatorname{supp} f\subseteq S(\delta_0)$. Since $S(\delta_0)$ is compact, there is an $\eta\in\mathcal{S}(\mathbb{R}^n)$ with $0\leq\eta\leq 1$, $\eta=1$ on $\operatorname{supp} f$, $\eta=0$ on the closure of $\mathbb{R}^n\setminus S(\delta_0/2)$. Pick $\{\phi_m\}\subseteq\mathcal{S}(\mathbb{R}^n)$ such that $\phi_m\to f$ in L^2 norm. Then $\eta\phi_m\to\eta f=f$ in L^2 norm and

$$\|B\eta\phi_l - B\eta\phi_m\|_\mu \leq \operatorname*{ess\,sup}_{\mathbf{x}\in S(\delta_0/2)} |\alpha(\mathbf{x})|\, d(S(\delta_0/2))\|\phi_l-\phi_m\|.$$

Since B is closed, $f\in D(B)$.

Corollary: If $A=\int\alpha(\mathbf{x})\,dE_\mathbf{x}$, $\Gamma(\alpha)$ of Lebesgue measure zero, and A has property P, D_0 is a core for A.

III. APPLICATIONS

If α is polynomially bounded (e.g., free Hamiltonians), $\Gamma(\alpha)$ is trivially of measure zero and property P follows from the recurrence relation for the h.o. functions. For $\alpha\in L^p(\mathbb{R}^n)$, $2\leq p\leq\infty$ (including many potentials such as the Yukawa potential), Hölder's inequality and the fact that there is a single constant[5] K such that $\|\psi_\mathbf{j}\|_\infty < K$ for all h.o. functions establishes property P. The condition on $\Gamma(\alpha)$ is usually obvious in each case.

External potentials are described by α's of the form $\alpha(\mathbf{y}) = \int V(\mathbf{y}-\mathbf{x})\,d\rho(\mathbf{x})$, ρ describing a fixed external source distribution with Radon–Nikodym derivative denoted ρ' and $V(\mathbf{y}-\mathbf{x})$ a two-body potential. Property P for several classes of these external potentials is established as follows:

Lemma: Let V be such that there is some $\beta = (\beta_1, \ldots, \beta_n)$, $\beta_i \geq 0$ for which $\|\Pi_{i=1}^n (1 + |x_i|)^{-\beta_i} V\|_\infty$ is finite and $\mathbf{x}^\alpha \in L_\rho^1(\mathbb{R}^n)$ for all $\alpha = (\alpha_1, \ldots, \alpha_n)$ such that $0 \leq \alpha_i \leq \beta_i$. Then $\alpha = V * \rho$ is polynomially bounded and hence has property P.

Proof: In the convolution integral defining α, multiply and divide by $\Pi_{i=1}^n (1 + |x_i - y_i|)^{\beta_i}$ and expand the numerator by the multinomial theorem and triangle inequality.

Lemma: Let $V \in L^p(\mathbb{R}^n)$, $2 \leq p \leq \infty$, ρ a finite positive measure. Then α has property P.

Proof: Integrating $V * \rho$ against an L^q function and interchanging the convolution shows $V * \rho \in L^p(\mathbb{R}^n)$ for $1 \leq p \leq \infty$.

Lemma: $V \in L^p(\mathbb{R}^n)$, $p > 1$, $\rho' \in L^q(\mathbb{R}^n)$, $p^{-1} + q^{-1} = 1$ implies $V * \rho \in L^\infty(\mathbb{R}^n)$ and hence has property P. (Use Hölder's inequality).

Lemma: If V is a Rollnik potential and ρ' is such that $|\mathbf{x}|^s \rho'(\mathbf{x}) \in L^\infty(\mathbb{R}^3)$ for $s = 0, 1, 2$, then $V * \rho' \in L^\infty(\mathbb{R}^3)$ and hence has property P. [For $|V * \rho'|^2(z)$, multiply and divide by $|\mathbf{x}-\mathbf{y}|^2$, and use the triangle inequality in the numerator.]

We remark that the roles of V, ρ in the above may be interchanged whenever V is locally integrable. Furthermore, if D denotes derivative and V is locally integrable, the relation $V * D^\alpha \rho = D^\alpha V * \rho$ may be useful in transfering singular behavior between V and ρ.

For bona fide two-body potentials, $A = \int V(\mathbf{x}-\mathbf{y}) \times d(E_\mathbf{x} \otimes E_\mathbf{y})$, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, so that $\Gamma(V)$ is trivially of measure zero, being confined to a hyperplane.

Lemma: For V polynomially bounded, or $V \in L^2(\mathbb{R}^n)$, $2 \leq p \leq \infty$, or $V =$ Fourier transform of Rollnik $\cap L^1(\mathbb{R}^3)$, then the two-body potential A has property P.

Proof: If V is such that there is a polynomial Q with $V^2 Q^{-1} \in L^1(\mathbb{R}^n)$ (V viewed as a function on $\mathbb{R}^n$), then V has property P. [In $\|A\psi_j\|^2$ multiply and divide by $Q(\mathbf{x}-\mathbf{y})$ and expand the numerator with the multinomial expansion. Then use the recurrence relations for the h.o. functions to obtain constants c' such that $\|\mathbf{x}^\beta \psi_j(\mathbf{x})\|_\infty \leq c'(1 + |j|)^{|\beta|}$.] For $V \in L^2(\mathbb{R}^n)$ we may choose $Q \equiv 1$; for $V \in L^\infty(\mathbb{R}^n)$, choose $Q^{-1} \in L^1(\mathbb{R}^n)$; for V polynomially bounded, choose Q to be a polynomial of sufficiently higher degree and with no real zeros; for $V =$ Fourier transform of Rollnik $\cap L^1(\mathbb{R}^3)$, use $Q(\mathbf{z}) = \mathbf{z}$; and finally if $V \in L^p(\mathbb{R}^n)$, $2 < p < \infty$, choose $Q^{-1} \in L^{p/p-2}(\mathbb{R}^n)$, and use $\int |V|^2 |Q^{-1}| \leq \| V^2 \|_{p/2} \| Q^{-1} \|_{p/p-2}$.

Slight modifications of the above results may be used when measures μ are introduced, as in the case of relativistic measures for various spin and space dimensions.

The above results, showing that under very general circumstances D_0 is a core for A, may be coupled with a known result[6] to conclude that the second-quantized form of A has a series expansion in terms of the free quantum fields in both the Fermi and Bose cases, thereby providing a generalization of a theorem of Hepp.[7]

*Supported in part by a grant from the Florida Atlantic University Division of Sponsored Research.

[1] F.E. Schroeck, Jr., J. Math. Phys. **14**, 130 (1973); **16**, 729 (1975).
[2] L. Schwartz, *Théorie des distributions* (Hermann, Paris, 1959), Vol. II.
[3] Ref. 2, pp. 117–19.
[4] E. Nelson, Ann. Math. **7**, 583 (1959).
[5] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions* (U.S. Government Printing Office, Washington, D.C., 1966), p. 787.
[6] F.E. Schroeck, Jr., J. Math. Phys. **12**, 1849 (1971).
[7] K. Hepp, *Théorie de la renormalization* (Springer, New York, 1969), pp. 24, 25.

An exact spherically symmetric local solution in a class of "strong" and "weak" two-tensor gravity theories

William C. Hammel

Department of Physics, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201
(Received 15 April 1974; revised manuscript received 7 June 1974)

After defining simultaneous spherical symmetry for two metrics, the technique of equation splitting is used to produce exact spherically symmetric solutions within the class of two-metric gravity theories introduced in an earlier paper. The nontrivial solutions are found to select the exceptional case, distinguished in previous work. The solutions in this case, however, possess a "gauge freedom" which apparently leaves the physics ambiguously defined. Nevertheless, we use the gauge freedom to simplify and proceed to local solutions, where ${}^1g_{..}$ is an empty closed space of positive constant curvature and ${}^0g_{..}$ is a Schwarzschild solution with a mass. The empty closed space is found to be contained within the Schwarzschild radius.

I. INTRODUCTION

In an earlier paper[1] Lubkin and the author have argued that the causal structure of space–time in the neighborhood of hadronic matter may be dominated by a massive spin-2 meson or rather by a fictitious meson which represents the average effect of the known spin-2 octet. We introduce cosmological terms as mass terms. The argument against such an identification in a single metric context does not hold[2] in the presence of two dynamical metrics. In Ref. 1 we introduced a general class of Lagrangian theories of two minimally coupled, second rank symmetric tensor fields that is based on the idea that all matter other than hadronic should be universally coupled to the "gravitational" field; in particular, the gravitation field is coupled to itself. Also all hadronic matter is to be coupled universally to the heavy spin-2 meson; and, again, in particular, the heavy meson is coupled to itself. The self-coupling is to be generated through the field energy, in the manner of Gupta.[3] The coupling between the two fields is prescribed by a generally covariant field-current identity and a criterion of minimality: that there be no derivatives and that the interaction Lagrangian density be the simplest possible function of the two tensor fields.

We point out that others have also seen the desirability of such 2-tensor gravity theories.[4] Since the motivation is, by now, a part of the literature, we refer the reader to the references. Our Lagrangian given in the next section is an alternative to those suggested by Isham, Salam, and Strathdee, and by Aichelburg and Mansouri; in fact, our Lagrangian includes that of Aichelburg and Mansouri as a subcase; the evidence so far is that our Lagrangian is a real generalization of theirs.

Although some effort has gone into investigating the nature of the solutions to above-mentioned 2-tensor theories,[5] the results have been somewhat disappointing. Mansouri and Urbantke have shown, for instance, that the Kerr–Schild ansatz

$$g_{\mu\nu}=\eta_{\mu\nu}-F(x)N_\mu N_\nu$$

where N_μ is null with respect to either $\eta_{..}$ or $g_{..}$, for both tensor fields will allow only pp-wave solutions: this, in spite of the well-known richness of the K–S ansatz in general relativity.

A solution notably missing from the 2-metric K–S family of solutions is a "bi-Schwarzschild" solution.

Moreover, an exact spherically symmetric static solution to either of the other theories has not yet been found. For our Lagrangian, we give a special, but exact spherically symmetric solution to the 2-tensor field equations. Although this solution is in a subcase of the Lagrangian that we previously[1,2] called the exceptional case, the solution may yet provide clues as to the behavior of more general solutions. It can be remarked that the exceptional subcase has no intersection with the Aichelburg–Mansouri subcase.

In Refs. 1,2 we have introduced a generally covariant system of equations for two "metric" tensors ${}^0g_{\mu\nu}$ and ${}^1g_{\mu\nu}$ that are derived from the Lagrangian density (upper and lower dot pairs indicate contractions):

$$\mathcal{L}=\frac{1}{{}^0k^2}\sqrt{-{}^0g}({}^0R-2{}^0\Lambda)+\frac{1}{{}^1k^2}\sqrt{-{}^1g}({}^1R-2{}^1\Lambda)$$
$$+\frac{1}{{}^0k^2}{}^0\mu\sqrt{-{}^0g}{}^1g_{..}{}^0g^{\cdot\cdot}+\frac{1}{{}^1k^2}{}^1\mu\sqrt{-{}^1g}{}^0g_{..}{}^1g^{\cdot\cdot}. \tag{1.1}$$

${}^0k$, ${}^1k$, ${}^0\Lambda$, ${}^1\Lambda$, ${}^0\mu$, ${}^1\mu$ are free parameters; we refer the reader to Refs. 1 and 2 or [6] for curvature conventions. By varying ${}^0g^{\mu\nu}$ and ${}^1g^{\mu\nu}$ independently, one obtains the field equations

$${}^0G_{\mu\nu}+({}^0\Lambda-\tfrac{1}{2}{}^0\mu{}^0g^{\cdot\cdot}{}^1g_{..}){}^0g_{\mu\nu}$$
$$={}^0k^2{}^1\mu\frac{\sqrt{-{}^1g}}{\sqrt{-{}^0g}}{}^1g^{\cdot\cdot}{}^0g_{\cdot\mu}{}^0g_{\cdot\nu}-{}^0\mu{}^1g_{\mu\nu}, \tag{1.2a}$$

$${}^1G_{\mu\nu}+({}^1\Lambda-\tfrac{1}{2}{}^1\mu{}^1g^{\cdot\cdot}{}^0g_{..}){}^1g_{\mu\nu}$$
$$={}^1k^2{}^0\mu\frac{\sqrt{-{}^0g}}{\sqrt{-{}^1g}}{}^0g^{\cdot\cdot}{}^1g_{\cdot\mu}{}^1g_{\cdot\nu}-{}^1\mu{}^0g_{\mu\nu} \tag{1.2b}$$

where

$$k\equiv{}^0k/{}^1k. \tag{1.3}$$

For details on constant solutions to (1.2), and other results we refer the reader to Refs. 1,2.

Although the original intent of the Eqs. (1.2) was to model the mixing of massless and massive spin-2 fields, it has become clear that the values of the 5 parameters ${}^0\Lambda$, ${}^1\Lambda$, ${}^0\mu$, ${}^1\mu$, k, which were introduced for generality,

influence greatly the nature of the solutions of the field equations: The nonperturbative linearization scheme introduced in Ref. 2 indicates seven distinct behavioral possibilities which are characterized by mass and spin of whatever particles remain in the reduction with respect to the Lorentz group into irreducible parts of a second rank symmetric tensor. In the analysis of constant solutions,[1] it has been shown that there exist four basic behavioral patterns.

In the analysis of constant solutions it was found convenient to introduce the matrix M, whose components are given by

$$M^i{}_j \equiv {}^0g^{i\cdot} {}^1g_{\cdot j} \tag{1.4}$$

The constant solutions could then be discussed in terms of Jordan-structure of M. Since for constant solutions, ${}^0G_{\mu\nu} = {}^1G_{\mu\nu} = 0$, M satisfies two simultaneous quadratic equations, the possible Jordan forms of M are thereby restricted, and it turns out that they are controlled by the values of the parameters. The results of these restrictions are classified in Theorems 5 and 6 in Ref. 1. This algebraic analysis is of course useful in considering solutions to (1.2) that are more general than constant; the analysis also provides the analysis of the algebraic structure at a point (as opposed to over the whole manifold) of ${}^0g_{\mu\nu}$ and ${}^1g_{\mu\nu}$ that may be obtained by an equation splitting which determines a differential and an algebraic problem. In Ref. 1 we have called a solution such that ${}^0G_{\mu\nu} = {}^1G_{\mu\nu} = 0$, *bi-Einstein*.

Remark: The attitude taken for the Eqs. (1.2) is that they are to be understood to be defined on a connected, Hausdorff C^∞ manifold. Interesting complications of the manifold structure can be expected if one requires that the manifold also be inextendible, for now the concept of maximal analytic extension can be dichotomic. The difficulty of "simultaneous" extension is in fact seen in the solution presented here. [See Sec. V.]

In this paper we will have recourse to results of the algebraic problem in M contained in Ref. 1 to produce an exact spherically symmetric solution to (1.2). The appropriate theorems constitute the following section.

II. EQUATION SPLITTING, SOLUTIONS WITH ${}^0g$ EINSTEIN-LIKE, AND RESTRICTIONS ON LOCAL STRUCTURE

A. The algebraic M problem

Equation splitting consists in splitting terms off a single equation which equates a partial differential expression to zero, and then equating the remaining expression and the split-off expression to zero. This yields two equations which are simpler to solve than the original. On obtaining solutions to each of the simpler equations, one then tries to adjust one of them so that it is also a solution of the other equation.

The two sets of partial differential equations (1.2) have many possible splittings, however, one of them recommends itself most highly. Not only are solutions of ${}^0G_{ij} = 0$ available, but if one of these solutions can be matched with the remaining equations, a solution to the general equations which most closely resembles ordinary gravity for ${}^0g_{ij}$ results.

If this splitting is used in the 0-equation of (1.2), the equation is split into a *differential problem* and an *algebraic problem*. The algebraic problem is essentially a quadratic equation for the matrix M, and is a restriction on its algebraic structure. This restriction is now explicitly investigated. Consider the splitting of the field equations (1.2) obtained by setting ${}^0G_{ij} = 0$, thus only the 0-field equations (1.2a) is split; the remaining terms of (1.2a) must therefore sum to zero yielding

$$\sqrt{-{}^0g}[{}^0\Lambda {}^0g_{ij} - \tfrac{1}{2}{}^0\mu({}^0g^{\cdot\cdot} {}^1g_{\cdot\cdot}){}^0g_{ij} + {}^0\mu {}^1g_{ij}] - \sqrt{-{}^1g}[k^2 {}^1\mu {}^1g^{kl} {}^0g_{ki} {}^0g_{lj}] = 0.$$

After contracting this equation with ${}^0g^{hi}$, it may be put in the form

$$[{}^0\Lambda - \tfrac{1}{2}{}^0\mu({}^0g^{\cdot\cdot} {}^1g_{\cdot\cdot})]\delta^h_j + {}^0\mu {}^0g^{hi} {}^1g_{ij} - \frac{\sqrt{-{}^1g}}{\sqrt{-{}^0g}} {}^1\mu k^2 {}^1g^{hi} {}^0g_{ij} = 0.$$

Using the definition (1.4) and suppressing indices this equation may be written in terms of the matrix M:

$$[{}^0\Lambda - \tfrac{1}{2}{}^0\mu(\mathrm{Tr}M)]I + {}^0\mu M - [\det M]^{1/2} {}^1\mu k^2 M^{-1} = 0$$

where I is the 4×4 unit matrix. Multiplication by M yields a quadratic equation (2.1) which with the equation splitting assumption is assumed to hold over some region of the manifold.

We solve the algebraic problem for the single splitting of Eq. (1.2a) in a series of steps; later the additional restrictions obtained by the "double splitting" of both (1.2a) and (1.2b) into algebraic and differential problems are considered. The steps of the single splitting are: First, the quardratic nature of the Eq. (2.1) which is to be satisfied by a 4×4 matrix M is a restriction on the Jordan form of M. The allowable Jordan forms are found; this step depends only on the quadratic nature of (2.1). Second, if these forms are substituted into (2.1) the situation arises that the form solves or does not solve (2.1) depending upon whether certain of the theory's parameters vanish or do not vanish, and, also upon whether or not they satisfy certain other algebraic constraints. In some cases the eigenvalues will be determined by (2.1) and in some cases not. We therefore distinguish separately the combination of vanishing and not vanishing of all the parameters, then in each of these solutions all the allowable Jordan forms with unspecified eigenvalues may be tested, whereupon any constraints on the parameters or determination of eigenvalues arise explicitly. If the double splitting assumption is superposed on the results of single splitting, the additional necessary and sufficient conditions are contained in a stated lemma. The application of the lemma then produces the double splitting conclusions in a straightforward way.

The algebraic problem states therefore, that M must satisfy the matrix equation

$$M^2 + \left(\frac{{}^0\Lambda}{{}^0\mu} - \tfrac{1}{2}\mathrm{Tr}M\right)M - k^2\frac{{}^1\mu}{{}^0\mu}(\det M)^{+1/2}I = 0. \tag{2.1}$$

Either (i) this is M's minimal polynomial, or (ii) M has only one characteristic value and is diagonal:

$${}^1g_{ij} = \alpha {}^0g_{ij}.$$

The *constant* value of α can be calculated from (2.1) and the result is

$$\alpha = \frac{{}^0\Lambda}{\mu} \quad \text{if } \alpha \neq 0 \tag{2.2}$$

where

$$\mu \equiv {}^0\mu + k^2\,{}^1\mu. \tag{2.3}$$

The Jordan forms are now given: consider *Case* (i). Let (2.1) be M's minimal polynomial. Being quadratic there are either

(i. a) two distinct real roots appearing diagonally

$$\begin{pmatrix} \lambda_1 & & & \\ & \lambda_1 & & \\ & & \lambda_2 & \\ & & & \lambda_2 \end{pmatrix}, \quad \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \lambda_2 & \\ & & & \lambda_2 \end{pmatrix}$$

or

(i.b) one real root appearing as

$$\begin{pmatrix} \lambda & & & \\ & \lambda & & \\ & & \lambda & 1 \\ & & 0 & \lambda \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \lambda & 1 & & \\ 0 & \lambda & & \\ & & \lambda & 1 \\ & & 0 & \lambda \end{pmatrix}$$

or

(i.c) a pair of complex conjugate roots each appearing twice.

This exhausts the algebraic possibilities of the local structure of solutions obtained via the indicated equation splitting.

The two major alternatives following Eq. (2.1) are, of course, the direct result of the quadratic nature of (2.1). In general, substitution into (2.1) of the Jordan forms just listed, will place restrictions on the characteristic values, and on the parameters of the theory. There is, moreover, an additional complication which results from certain of the parameters possibly having the value zero; this possibility allows modification of the constraining power of (1.2). To be certain of exhaust ing the possible behaviors of M, as determined by (1.2), the Jordan canonical behavior will now be exposed taking into account these various complications; the actual results of the analysis and not the tedious details are contained here.

Theorem: All possible Jordan canonical forms of the matrix M constrained by Eq. (1.2) can be given as follows:

(I) If ${}^0\Lambda \neq 0$, ${}^0\mu \neq 0$, ${}^1\mu \neq 0$, there are only two possible structures:

1. $M = \lambda I$, $\lambda = {}^0\Lambda/\mu$, $\mu \neq 0$

or

2. $M = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \lambda_2 & \\ & & & \lambda_2 \end{pmatrix}$,

$\lambda_1 = v^2\lambda_2$, ${}^0\mu^2 \neq {}^1\mu^2 k^4$, i.e., $v^2 \neq 1$,

$\lambda_2 = 2\,{}^0\Lambda/{}^0\mu(1 - v^2)$

where $v \equiv -({}^1\mu/{}^0\mu)k^2$ with $v > 0$.

(II) If ${}^0\Lambda = 0$, ${}^0\mu \neq 0$, ${}^1\mu \neq 0$ there are five structures to be distinguished:

1. $M = \lambda I$, $\mu = 0$

and λ is real and not zero but otherwise undetermined;

2. $M = \begin{pmatrix} \lambda & 1 & & \\ 0 & \lambda & & \\ & & \lambda & 0 \\ & & 0 & \lambda \end{pmatrix}$ or $\begin{pmatrix} \lambda & 1 & & \\ 0 & \lambda & & \\ & & \lambda & 1 \\ & & 0 & \lambda \end{pmatrix}$, $\mu = 0$

and λ is real and not zero but otherwise undetermined;

3. $M = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_1 & & \\ & & \lambda_2 & \\ & & & \lambda_2 \end{pmatrix}$, $\lambda_1 \neq -\lambda_2$,

λ_1, λ_2 are otherwise undetermined if real, but are complex conjugates if complex, and

$$\lambda_1\lambda_2 > 0 \quad \text{if } {}^0\mu + k^2\,{}^1\mu = 0,$$

but

$$\lambda_1\lambda_2 < 0 \quad \text{if } {}^0\mu - k^2\,{}^1\mu = 0.$$

(III) If ${}^0\Lambda \neq 0$, ${}^0\mu \neq 0$, ${}^1\mu = 0$ there is only one structure possible:

$M = \lambda I$, $\lambda = {}^0\Lambda/{}^0\mu$, $\mu \neq 0$.

(IV) If ${}^0\Lambda \neq 0$, ${}^0\mu = 0$, ${}^1\mu \neq 0$ there is only one structure possible:

$M = \lambda I$, $\lambda = k^{-2}\,{}^0\Lambda/{}^1\mu$, $\mu = 0$.

Note: Cases (III) and (IV) are obviously boundary cases in parameter space of case (I.1.).

(V) If ${}^0\Lambda = 0$, ${}^0\mu \neq 0$, ${}^1\mu = 0$ then $M = 0$; therefore we drop the case.

(VI) If ${}^0\Lambda \neq 0$, ${}^0\mu = 0$, ${}^1\mu = 0$ then either $M = 0$ or $\mathrm{Tr}M = 0$, but M is otherwise undetermined and there are no further constraints on the parameters.

(VII) If ${}^0\Lambda = 0$, ${}^0\mu = 0$, ${}^1\mu \neq 0$ M must be singular; therefore we drop this case.

So from the viewpoint of two g-cones per point, solutions which may be obtained by the simplifying assumption of single equation splitting are not necessarily trivial, i.e., monoconic (${}^0g_{..} = \alpha\,{}^1g_{..}$).

B. Double equation splitting

In this subsection the concept of equation splitting is pushed further.

If in addition to ${}^0G_{ij} = 0$, ${}^1G_{ij} = 0$ is also imposed, M must satisfy an additional matrix equation, which by the nature of the full equation set is the 0 – 1 dual of (1.2); this equation may then be gotten formally by exchanging

0 and 1 in (1.2); this exchanges $k \to k^{-1}$, $M \to M^{-1}$. The result is

$$M^2 - \frac{{}^1\mu}{{}^0\mu} k^2 (\det M)^{1/2} \left(\frac{{}^1\Lambda}{{}^1\mu} - \tfrac{1}{2} \operatorname{Tr} M^{-1} \right)$$

$$\times M - k^2 \frac{{}^1\mu}{{}^0\mu} (\det M)^{1/2} I = 0. \tag{2.4}$$

The additional information contained in (2.4) beyond that in (2.1) is given by the following lemma:

Lemma: If M satisfies (2.1) the necessary and sufficient conditions that it also satisfies (2.4) are given by

(A) $(\det M)^{1/2} = k^{-2}({}^0\Lambda / {}^1\Lambda)$,

(B) $4\,{}^0\Lambda\,{}^1\Lambda = {}^0\mu\,{}^1 \operatorname{Tr} M + {}^1\mu\,{}^0 \operatorname{Tr} M^{-1}$.

Proof: To prove the necessity of (A) multiply (2.1) and (2.4) by M^{-1}, take determinants and solve for $(\det M)^{1/2}$ To prove the necessity of (B), subtract (2.1) and (2.4) and use (A). Sufficiency is mere substitution.

If one applies the foregoing lemma to each of the cases in the previous theorem bearing in mind that both ${}^1\Lambda = 0$, ${}^1\Lambda \neq 0$ must be considered, the surviving cases (the others demand singular M), are given by the following theorem.

Theorem: The nonsingular Jordan canonical forms of all M constrained by Eqs. (2.1) and (2.4) can be given as follows: We list classification numbers from the previous theorem only where they survive and state only the *additional* information supplied by (2.4) or equivalently by the preceeding lemma.

(I) 1. ${}^0\Lambda\,{}^1\Lambda = (k^{-1}\,{}^0\mu + k^1\mu)^2 \equiv \tilde{\mu}^2$, ${}^1\Lambda \neq 0$,

2. $4 \frac{{}^0\Lambda\,{}^1\Lambda}{{}^0\mu\,{}^1\mu} v^2 = -(1 - v^2)^2$, $v \equiv -\frac{{}^1\mu}{{}^0\mu} k^2$;

(II) ${}^1\Lambda = 0$ but no additional information;

(III) ${}^0\Lambda\,{}^1\Lambda = k^{-2}\,{}^0\mu^2$, ${}^1\Lambda \neq 0$;

(IV) ${}^0\Lambda\,{}^1\Lambda = k^2\,{}^1\mu^2$, ${}^1\Lambda \neq 0$.

III. SPHERICAL SYMMETRY

A. Spherical symmetry in a single metric context

Let $g_{\mu\nu}$ be a sufficiently differentiable symmetric tensor field with hyperbolic signature on a smooth manifold (any singular points are not part of the manifold if one requires local Minkowskian structure for the manifold); $g_{\mu\nu}$ is said to be spherically symmetric if there exists a set of three space-like vectors v_a^μ, $a = 1,2,3$, $\mu = 1,2,3,4$ such that

$$\underset{c}{\pounds}_b v^\mu = C^a_{cb} v^\mu_a \quad \text{(Ref. 7)} \tag{3.1}$$

where C^a_{cb} are the structure constants of the Lie group $SO(3)$ and $\pounds_c$ is the Lie derivative operator with respect to v_c, and

$$\underset{a}{\pounds}\, g_{\mu\nu} = 0. \tag{3.2}$$

The manifold action of the isometry group $SO(3)$ is also taken to be transitive on orbits of space-like 2-surfaces, which are in fact 2-surfaces of constant curvature.[8]

B. Spherical symmetry in a two-"metric" context

In extending the concept of spherical symmetry to two hyperbolic metrics, attention must be paid to the requirement that the Killing vectors v^μ be space-like, for now there are two such concepts. If the two cones defined by ${}^0g_{\cdot\cdot}$ and ${}^1g_{\cdot\cdot}$ coalesce there is no problem; biconicity can be classified in two ways: the cones are concentric but distinct or their symmetry axes are distinct. These two situations engender two possible definitions of simultatneous spherical symmetry:

Strong spherical symmetry is said to hold for ${}^0g_{\cdot\cdot}$ and ${}^1g_{\cdot\cdot}$ if there exists a single set of Killing vectors obeying (3.1) which are space-like for both ${}^0g_{\cdot\cdot}$ and ${}^1g_{\cdot\cdot}$, and these Killing vectors generate transitive actions on two-surfaces space-like with respect to both ${}^0g_{\cdot\cdot}$ and ${}^1g_{\cdot\cdot}$.

Weak spherical symmetry is said to hold for ${}^0g_{\cdot\cdot}$ and ${}^1g_{\cdot\cdot}$ if there exists a set of 0-*Killing* vectors, obeying (3.1), whose elements are 0-*space-like*, and there exists another set of vectors obeying (3.1) all of whose elements are 1-Killing and 1-*space-like*. Moreover, each set of Killing vectors generate transitive actions on their respective space-like 2-surfaces.

In the context of strong spherical symmetry the still stronger condition of *strong spheristasis* can be imposed in exact analogy to the single-metric case; we have occasion to use this in Sec. IV.

Weak spherical symmetry is weak enough to include the bizarre possibilty that the manifold coordinates do not have the same physical interpretation for both tensor fields. This already occurs, for different manifold regions, in GR in the R- and T-regions of the Schwarzschild solution.[9] We note that the solution which we present does not present such pathology.

In the following we will consider a restricted form of weak spherical symmetry. Even so, this is still sufficiently weak a concept that our method for obtaining the solution leaves one functional degree of freedom undetermined. We discuss this situation in the final section.

If $g_{\cdot\cdot}$ is spherically symmetric and given in terms of spherical polar coordinates, and a time variable t, it has the well-known form:

$$ds^2 = \exp[\beta(r,t)dt^2] - [\exp(\alpha(r,t))\,dr^2 + \rho(r,t)\,d\Omega^2 + 2\gamma(r,t)dr\,dt] \tag{3.3}$$

where

$$d\Omega^2 \equiv d\theta^2 + \sin^2\theta\, d\phi^2; \tag{3.4}$$

the center of symmetry is of course $r = 0$.

Coordinate transformations[10]

$$r' = f(r,t), \quad t' = \tau(r,t), \quad \theta' = \pm\,\theta + N\pi, \quad \phi' = \phi + M\pi \tag{3.5}$$

where M, N are integers, and f and τ are arbitrary functions of r,t such that

$$f_r \tau_t - f_t \tau_r \neq 0, \tag{3.6}$$

preserve the Killing vectors of (3.3). (The subscripts r,t on functions indicate partial derivatives.)

Since our solution method will force the hypothesis of Birkhoff's theorem we define:

$g_{\cdot\cdot}$ is *static* if $g_{\cdot\cdot}$ has a time-like hypersurface orthogonal Killing vector. Thus, there exists a coordinate system such that

$$\frac{\partial}{\partial t} g_{\mu\nu} = 0. \tag{3.7}$$

Restricting the $g_{\cdot\cdot}$ of (3.3) also to be static, limits the transformations (3.5) further to

$$\begin{aligned} &r' = f(r), \quad t' = t + h(r), \\ &\theta' = \pm\,\theta + N\pi, \quad \phi' = \phi + M\pi. \end{aligned} \tag{3.8}$$

The form (3.3) is then also restricted so it can always be cast in the form

$$ds^2 = \exp[\beta(r)]dt^2 - (\exp[\alpha(r)]dr^2 + r^2\,d\Omega^2). \tag{3.9}$$

Both ${}^0g_{\cdot\cdot}$ and ${}^1g_{\cdot\cdot}$ may be put in the form (3.9) but in general not simultaneously. Let the unprimed coordinate system be that for which ${}^0g_{\cdot\cdot}$ assumes the form (3.9) and and the primed system be the one for ${}^1g_{\cdot\cdot}$. In the next section we exploit the existence of these two "simplifying" coordinate systems to obtain the solution.

IV. THE SOLUTION

In Sec. II a technique of equation splitting was introduced for Eq. (1.2) which determined a differential equation problem, an algebraic problem, and a matching problem. Equations (1.2) are written in the form

$${}^0G_{\mu\nu} + {}^0\Lambda\,{}^0g_{\mu\nu} = {}^0J_{\mu\nu}, \tag{4.1a}$$

$${}^1G_{\mu\nu} + {}^1\Lambda\,{}^1g_{\mu\nu} = {}^1J_{\mu\nu}. \tag{4.1b}$$

The equations are then split giving

$${}^0G_{\mu\nu} + {}^0\Lambda\,{}^0g_{\mu\nu} = 0, \tag{4.2a}$$

$${}^1G_{\mu\nu} + {}^1\Lambda\,{}^1g_{\mu\nu} = 0, \tag{4.2b}$$

$${}^0J_{\mu\nu} = 0, \tag{4.3a}$$

$${}^1J_{\mu\nu} = 0. \tag{4.3b}$$

The solution to the two parts of the differential problem (4.2) in respectively simplifying coordinates is well known when spheristasis is imposed. In fact Birkhoff's theorem shows that only static solutions are possible for spherically symmetric metrics satisfying vacuum Einstein equations.[11]

Also, when spheristasis holds, the algebraic problem posed by (4.3) is almost solved: (4.3) represents two quadratic equations in the matrix M defined by (1.4). Spheristasis leaves M almost in its Jordan-form, and the allowable Jordan-forms have been completely analyzed in the theorems of Sec. II. The matching problem is the result of fitting our calculated M to one of the allowed forms. In doing this, M must be calculated entirely in one (unprimed) coordinate system. ${}^1g_{\cdot\cdot}$ is given in primed coordinates; we then map back to unprimed coordinates using *a priori* unknown functions from (3.8) (we drop the transformations of angular variables), which are, subsequently, to be determined by (3.3) and the algebraic theorems of Sec. II.

Note: We might have considered the more general splitting

$$\begin{aligned} &{}^0G_{\mu\nu} + {}^0\lambda_1\,{}^0g_{\mu\nu} = 0, \quad {}^0J_{\mu\nu} - {}^0\lambda_2\,{}^0g_{\mu\nu} = 0, \\ &{}^1G_{\mu\nu} + {}^1\lambda_1\,{}^1g_{\mu\nu} = 0, \quad {}^1J_{\mu\nu} - {}^1\lambda_2\,{}^1g_{\mu\nu} = 0, \end{aligned}$$

where

$${}^0\Lambda = {}^0\lambda_1 + {}^0\lambda_2, \quad {}^1\Lambda = {}^1\lambda_1 + {}^1\lambda_2.$$

The algebraic problem for spheristasis however, would then, force ${}^0\lambda_2 = {}^1\lambda_2 = 0$, so no generality is in fact gained.

Now (4.2) determines

$$[{}^0g_{\mu\nu}(x)] = \begin{bmatrix} +{}^0\gamma(r) & & & \\ & -{}^0\gamma^{-1}(r) & & 0 \\ & & -r^2 & \\ 0 & & & -r^2\sin^2\theta \end{bmatrix} \tag{4.4a}$$

$$[{}^1g'_{\mu\nu}(x')] = \begin{bmatrix} +{}^1\gamma(r') & & & \\ & -{}^1\gamma^{-1}(r') & & 0 \\ & & -(r')^2 & \\ 0 & & & -(r')^2\sin^2\theta \end{bmatrix} \tag{4.4b}$$

where for brevity we define

$${}^0\gamma(r) \equiv 1 - \frac{2^0m}{r} + \tfrac{1}{3}{}^0\Lambda r^2, \tag{4.5a}$$

$${}^1\gamma(r') \equiv 1 - \frac{2^1m}{r'} + \tfrac{1}{3}{}^1\Lambda (r')^2. \tag{4.5b}$$

The unknown coordinate transformation

$$r' = f(r), \quad t' = t + h(r) \tag{4.6}$$

gives the Jacobian matrix

$$\left[\frac{\partial x'^{\sigma}}{\partial x^{\lambda}}\right] = \begin{bmatrix} 1 & h_r & 0 & 0 \\ 0 & f_r & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{4.7}$$

So, the tensor transformation law gives ${}^1g_{\cdot\cdot}$ in unprimed coordinates:

$$[{}^1g_{\cdot\cdot}] = \begin{bmatrix} {}^1B & {}^1C & 0 & 0 \\ {}^1C & -{}^1A & 0 & 0 \\ 0 & 0 & -f^2 & 0 \\ 0 & 0 & 0 & -f^2\sin^2\theta \end{bmatrix}, \tag{4.8}$$

where

$$\begin{aligned} &{}^1B \equiv {}^1\gamma[f(r)] - \frac{h_r^2}{{}^1\gamma[f(r)]}, \\ &{}^1A \equiv \frac{f_r^2}{{}^1\gamma[f(r)]}, \\ &{}^1C \equiv -\frac{f_r h_r}{{}^1\gamma[f(r)]}. \end{aligned} \tag{4.9}$$

For notational convenience we redefine

$${}^0\gamma(r) = {}^0B(r). \tag{4.10}$$

Then, M defined by (1.4) is calculated in unprimed coordinates to be

$$M=\left[\begin{array}{cc|cc} \frac{{}^1B}{{}^0B} & \frac{{}^1C}{{}^0B} & & \\ -{}^0B^1C & -{}^1A^0B & \multicolumn{2}{c}{0} \\ \hline & & \left(\frac{f}{r}\right)^2 & \\ \multicolumn{2}{c|}{0} & & \left(\frac{f}{r}\right)^2 \end{array}\right] \tag{4.11}$$

The algebraic problem (3.3) determines two quadratic equations for M which can written[1] as in Sec. II:

$$M^2-(\tfrac{1}{2}\,\mathrm{Tr}M)M-k^2\frac{{}^1\mu}{{}^0\mu}(\det M)^{1/2}I=0, \tag{4.12a}$$

$$M^2+k^2\frac{{}^1\mu}{{}^0\mu}(\det M)^{1/2}(\tfrac{1}{2}\,\mathrm{Tr}M^{-1})M-k^2\frac{{}^1\mu}{{}^0\mu}(\det M)^{1/2}I=0. \tag{4.12b}$$

From (4.12) M can have at most *two* characteristic values, while (4.11) apparently allows *three*. So the algebraic problem will be a statement of degeneracy for the characteristic values of (4.11).

The two characteristic values of the upper 2×2 submatrix are

$$2S_{\mp}=\frac{{}^1B}{{}^0B}-{}^1A^0B\pm\left[\left(\frac{{}^1B}{{}^0B}+{}^1A^0B\right)^2-4{}^1C^2\right]^{1/2}. \tag{4.13}$$

From the algebraic problem as analyzed in the theorems of Sec. II we know that there are two classes of Jordan-forms for M, those where the characteristic values are determined functions of the parameters ${}^0\Lambda$, ${}^1\Lambda$, ${}^0\mu$, ${}^1\mu$, k^2 and therefore constant over the underlying manifold, and those where the characteristic values are not determined but yet we have algebraic constraints on the parameters. A situation where M is determined to be constant is clearly unsatisfactory as it can only lead to trivial solutions for our situation since it forces $f(r)=(\mathrm{const})r$.

After one abandons the trivial cases, only two remain:

$$(1)\quad M=\left(\frac{f(r)}{r}\right)^2 I, \tag{4.14}$$

where

$$S_+=S_-=\left(\frac{f(r)}{r}\right)^2. \tag{4.15}$$

For this we learn from Sec. II that $\mu\equiv{}^0\mu+k^2{}^1\mu=0$ must also hold. This the *exceptional case*.

$$(2)\quad M=\begin{bmatrix} S & 0 & & \\ 0 & S & \multicolumn{2}{c}{0} \\ & & \left(\frac{f}{r}\right)^2 & \\ \multicolumn{2}{c}{0} & & \left(\frac{f}{r}\right)^2 \end{bmatrix} \tag{4.16}$$

where

$$S\equiv S_+=S_-. \tag{4.17}$$

From the first theorem of Sec. II (Case II.3) we learn that $S\neq-(f/r)^2$ and that S, $(f/r)^2$ are undetermined by (4.12) if they are real but conjugates if they are complex and further, if they are real, that either

$${}^0\mu+k^2{}^1\mu=0\rightarrow S\left(\frac{f}{r}\right)^2>0 \tag{4.18}$$

$${}^0\mu-k^2{}^1\mu=0\rightarrow S\left(\frac{f}{r}\right)^2<0. \tag{4.19}$$

We return to these inequalities and their satisfaction later.

In both cases the discriminant in (4.13) vanishes:

$$\left(\frac{{}^1B}{{}^0B}+{}^1A^0B\right)^2=4{}^1C^2. \tag{4.20}$$

Using (4.9), (4.20) determines a nonlinear differential equation in the two unknown functions h and f.

$$[{}^1\gamma[f(r)]]^2-h_r^2+f_r^2{}^0\gamma^2(r)=4{}^0\gamma f_r^2h_r^2{}^1\gamma^{-1}. \tag{4.21}$$

The conformal case (1) will provide an additional differential equation for the two unknown functions h, and f; this case, although more complicated, may answer the question of the existence of conformally related solutions in the exceptional case.[12] We hope to return to this question in a future publication.

Before continuing to a solution of (4.21) it is worth remarking on the perhaps disturbing absence of control on the functions implied by (4.21); note also that in spite of the fact that either (4.18) or (4.19) is forced, μ's will not appear in the solution, thus their role as coupling constants is vitiated in these solutions. The equation splitting assumption has negated a good deal of the dynamical control exerted by the fields ${}^0g_{..}$ and ${}^1g_{..}$ on each other since they both determine vacuum Einstein spaces in the usual sense of GR. From this point of view one should not criticize the lack of control in (4.21) but marvel that it has any content; the significance of the freedom is, however, unclear at present.

Now, as long as the freedom is available, set

$$h(r)\equiv0. \tag{4.22}$$

(We are now in the regime of strong spheristasis.) Then (4.21) reduces to

$$[{}^1\gamma[f(r)]]^2+f_r^2{}^0\gamma^2(r)=0, \tag{4.23}$$

which in turn yields a linear, separable ordinary differential equation

$$\frac{df}{1-2{}^1m/f-[{}^1\Lambda(-)/3]f^2}=\pm i\frac{dr}{1-2{}^0m/r-[{}^0\Lambda(-)/3]r^2}, \tag{4.24}$$

where $i\equiv\sqrt{-1}$, and for calculational ease we define

$${}^0\Lambda(-)\equiv-{}^0\Lambda,\quad {}^1\Lambda(-)\equiv-{}^1\Lambda.$$

Although the integral to be performed in (3.24) can be done in principle [$\int ds/(1-a/s+bs^2)$ is a linear combination of three integrals of the form $\int s\,ds/(s+A)$], the work can be simplified by asking the following question:

Is it possible that a lump of pure ${}^1g$-curved space may give rise to a gravitational mass in ${}^0g$? As we will see the answer is yes and, moreover, the only region where deviations from GR are predicted for these solutions is the inside of Schwarzschild singular surface.

Setting the conditions for the above question,

$$ {}^1m \equiv 0, \quad {}^0\Lambda \equiv 0, \tag{4.25} $$

(4.24) then becomes

$$ \pm \frac{df}{1-[{}^1\Lambda(-)/3]f^2} = i\frac{dr}{1-2{}^0m/r}. \tag{4.26} $$

The $\pm$ distinction turns out to be useless so + is chosen uniformly from here on.

We will need to distinguish the cases ${}^1\Lambda > 0$, ${}^1\Lambda < 0$: If ${}^1\Lambda > 0$, with the signature chosen $(+---)$, this is the condition for a hyperbolic universe of constant negative curvature. (cf., however, the end of this section.)

Let ${}^1\Lambda > 0$; (4.26) integrates to

$$ \tan^{-1}|\sqrt{\Lambda/3}f| = \sqrt{\Lambda/3}[z + i(r - 2m + 2m\ln(r-2m))] \tag{4.27} $$

where z is a complex constant of integration.

Since there can be no confusion now, the left superscripts are to be dropped on ${}^1\Lambda$, and ${}^0m$.

Equation (4.27) can be rewritten as

$$ \sqrt{\Lambda/3}f = i\tanh[\sqrt{\Lambda/3}(-iz - 2m + r + 2m\ln(r-2m))]. \tag{4.28} $$

The coordinate transformation function $f(r)$ is in general a complex function. It will be convenient to define the complex function

$$ \alpha(r) \equiv \sqrt{\Lambda/3}(-iz - 2m + r + 2m\ln(r-2m)), \tag{4.29} $$

so

$$ \sqrt{\Lambda/3}f = i\tanh\alpha(r). \tag{4.30} $$

If $\Lambda < 0$ is chosen, then one merely replaces $\sqrt{\Lambda/3}$ with $\sqrt{-\Lambda/3}$ and tanh with tan in (4.29) and (4.30). It is clear from its functional form that f can be a legitimate coordinate transformation either inside or outside of the Schwarzschild radius.

Moreover, we must insure satisfaction of one of the inequalities (4.18) or (4.19). It is not hard to verify that for both signs of Λ

$$ S < 0 \text{ when } r < 2m \text{ and } \therefore\ {}^0\mu + k^2{}^1\mu = 0 \text{ must hold,} $$

and

$$ S > 0 \text{ when } r > 2m \text{ and } \therefore\ {}^0\mu - k^2{}^1\mu = 0 \text{ must hold;} $$

provided that $f(r)$ is purely imaginary.

The most physical thing to do seems not to interfere with GR outside of the Schwarzschild radius. So we restrict the domain of $f(r)$ to region $r < 2m$ and enforce the exceptional case $\mu \equiv {}^0\mu + k^2{}^1\mu = 0$. Then, we rewrite

$$ \alpha(r) = \sqrt{\Lambda/3}(-iz - 2m + r + 2m\ln(2m-r) + i2m\pi), \tag{4.31} $$

so

$$ \sqrt{\Lambda/3}f(r) = i\tanh\alpha(r) \quad \text{for } r < 2m, \tag{4.32} $$

and

$$ {}^1\gamma(r) = \operatorname{sech}^2\alpha(r). \tag{4.33} $$

Now in order to make f purely imaginary, we note that $\mathrm{Im}\alpha(r) = \sqrt{\Lambda/3}(2m\pi - z)$ is constant and further that $\tanh(z + iN\pi) = \tanh z$ for any integer N. Therefore, choosing any z_N, such that

$$ \sqrt{\Lambda/3}(2m\pi - z_N) = N\pi \tag{4.34}[13] $$

as constant of integration, is a general, sufficient condition for f to be purely imaginary. (If f were not so then the matrix M would be complex and the algebraic problem would trivialize the solution; this is a consequence of the first theorem of Sec. II.)

With the choice (4.34), $\alpha(r)$ becomes

$$ \alpha(r) = \sqrt{\Lambda/3}(r - 2m + 2m\ln(2m-r)) + iN\pi \tag{4.35} $$

and

$$ \sqrt{\Lambda/3}f(r) = i\tanh\sqrt{\Lambda/3}(r - 2m + 2m\ln(2m-r)). \tag{4.36} $$

Now since only $[f]^2$ enters the expressions for the components of ${}^1g_{\cdot\cdot}$ there will be no imaginary parts entering the final solution [this does not of course necessarily follow if (4.25) is not chosen]; the factor i will, however, tamper with the signature of ${}^1g_{\cdot\cdot}$:

$$ \begin{aligned} {}^1g_{\cdot\cdot}\,dx^\cdot\,dx^\cdot &= [\operatorname{sech}\alpha(r)]^2\,dt^2 - f_r^2[\operatorname{sech}\alpha(r)]^{-2}\,dr^2 \\ &\quad + \frac{3}{\Lambda}[\tanh\alpha(r)]^2\,d\Omega^2 \quad \text{for } r < 2m \end{aligned} \tag{4.37} $$

and, of course,

$$ {}^0g_{\cdot\cdot}\,dx^\cdot\,dx^\cdot = -\left(\frac{2m}{r} - 1\right)dt^2 + \left(\frac{2m}{r} - 1\right)^{-1}dr^2 - r^2\,d\Omega^2 \quad \text{for } r < 2m \tag{4.38} $$

We notice, however, that (4.23) makes

$$ [f_r]^2 = -{}^1\gamma^2/{}^0\gamma^2 < 0 $$

for both $r \gtrless 2m$ and $\Lambda \gtrless 0$.

Also using (4.33) we have, for $\Lambda > 0$ and $r < 2m$,

$$ {}^0g_{\cdot\cdot}\,dx^\cdot\,dx^\cdot = [{}^0\gamma(-)]^{-1}\,dr^2 - {}^0\gamma(-)dt^2 - r^2\,d\Omega^2 \tag{4.39a} $$

and

$$ {}^1g_{\cdot\cdot}\,dx^\cdot\,dx^\cdot = \frac{{}^1\gamma}{[{}^0\gamma(-)]^2}\,dr^2 + {}^1\gamma\,dt^2 + \frac{3}{\Lambda}(1 - {}^1\gamma)d\Omega^2. \tag{4.39b} $$

Notice that in the region $r < 2m$ ${}^0g_{\cdot\cdot}$ is hyperbolic and that the distinguished time variable must be chosen to be r; we choose r also as time variable for ${}^1g$ but ${}^1g$ happens to be positive definite, and so with $\Lambda > 0$ it determines a closed space of positive definite curvature. t is now a space-like variable and although we still have a symmetry for (4.39) from the fact that there obviously exists a coordinate system in which

$$ \frac{\partial}{\partial t}{}^0g_{\mu\nu} = \frac{\partial}{\partial t}{}^1g_{\mu\nu} = 0, \tag{4.40} $$

we may no longer call this symmetry stasis, since the Killing vector is now space-like with respect to ${}^0g_{\cdot\cdot}$. It is still however hypersurface orthogonal, with the hypersurface being time-like with respect to ${}^0g_{\cdot\cdot}$. It is interesting to note that in the region $r < 2m$ an analytic continuation[14] from $t \to it$ defines a hyperbolic ${}^1g$ where t is once again the distinguished time variable, while ${}^0g$

has signature $(+\,+\,-\,-)$. So on enlarging the space–time to the direct sum of two manifolds I and II with $r < 2m$ we have ${}^0g_{\cdot\cdot}$ and ${}^1g_{\cdot\cdot}$ both real and in

I: ${}^0g$, hyperbolic with time variable r and

${}^1g$, positive definite,

while in

II: ${}^1g$ hyperbolic with time variable t and

${}^0g$ with signature $(+\,+\,-\,-)$.

Is it possible to push this solution out of the region $r < 2m$? Considering only ${}^0g_{\cdot\cdot}$, the standard methods of null coordinates[8] in the manner of Eddington,[15] Finkelstein,[16] and Kruskal[17] can of course be used to unite the two manifolds $r < 2m$ and $r > 2m$. But what of ${}^1g_{\cdot\cdot}$?

Under the coordinate substitution

$$t \to v \equiv t + r^* \tag{4.41}$$

where

$$r^* \equiv r + 2m \ln(r - 2m), \tag{4.42}$$

Eq. (4.39b) becomes

$${}^1g_{\cdot\cdot}\, dx^{\cdot}\, dx^{\cdot} = {}^1\gamma\, dv^2 + 2\frac{{}^1\gamma}{{}^0\gamma^2}\, dr^2 - 2\frac{{}^1\gamma}{{}^0\gamma}\, dr\, dv + \frac{3}{\Lambda}(1 - {}^1\gamma)d\Omega^2, \tag{4.43a}$$

while it is well known that (4.39a) becomes

$${}^0g_{\cdot\cdot}\, dx^{\cdot}\, dx^{\cdot} = -\left(\frac{2m}{r} - 1\right) dv^2 - 2\, dv\, dr - r^2\, d\Omega^2. \tag{4.43b}$$

${}^0g_{\cdot\cdot}$ is nonsingular on the Schwarzschild surface, but the limit of ${}^1g_{\cdot\cdot}$ as $r \to 2m$ must be treated with care: The following limits are instantly available as $r \to 2m$:

$$\begin{aligned} &\gamma(r) \to -\infty, \quad f(r) \to i\sqrt{3/\Lambda}, \\ &{}^0\gamma(r) \to 0, \quad {}^1\gamma(r) \to 0. \end{aligned} \tag{4.44}$$

Also,

$$\lim_{r \to 2m} \frac{{}^1\gamma}{{}^0\gamma^2} = \begin{cases} 0 & \text{if } \sqrt{\Lambda/3} > 1/2m \\ 16m^2 & \text{if } \sqrt{\Lambda/3} = 1/2m \\ \infty & \text{if } \sqrt{\Lambda/3} < 1/2m \end{cases} \tag{4.45}$$

and

$$\lim_{r \to 2m} \frac{{}^1\gamma}{{}^0\gamma} = \begin{cases} 0 & \text{if } \sqrt{\Lambda/3} > 1/4m \\ -8m & \text{if } \sqrt{\Lambda/3} = 1/4m. \\ \infty & \text{if } \sqrt{\Lambda/3} < 1/4m \end{cases} \tag{4.46}$$

Therefore, if

$$\sqrt{\Lambda/3} > 1/2m, \tag{4.47}$$

$$\lim_{r \to 2m} {}^1g_{\cdot\cdot}\, dx^{\cdot}\, dx^{\cdot} = \frac{3}{\Lambda}\, d\Omega^2, \tag{4.48}$$

the same limit obtains in Schwarzschild coordinates. A more interesting limit where ${}^1g$ is "spatially nonsingular" occurs when $\sqrt{\Lambda/3} = 1/2m$, so that $\sqrt{\Lambda/3} > 1/4m$. Then,

$$\lim_{r \to 2m} {}^1g\, dx^{\cdot}\, dx^{\cdot} = 32m^2\, dr^2 + \frac{3}{\Lambda}\, d\Omega^2 = 4m^2(8\, dr^2 + d\Omega^2). \tag{4.49}$$

The limit (4.48) reminds us, however, of the bottle picture of an almost closed internal space which necks on to macroscopic space–time; only the angular part of ${}^1g_{\cdot\cdot}$ survives in the limit $r \to 2m$, so ${}^1g_{\cdot\cdot}$ almost vanishes. Now one may satisfy the field equation (1.2) in the region $r > 2m$ with the exterior Schwarzschild solution and

$${}^1g_{\cdot\cdot} \equiv 0 \quad \text{for } r > 2m. \tag{4.50}$$

Although ${}^0g_{\cdot\cdot}$ passes continuously between $r > 2m$ and $r < 2m$, the interior and exterior solutions for ${}^1g_{\cdot\cdot}$ suffer a finite saltus in the angular terms at $r = 2m$. Consequently, the field equations (1.2) cannot be rigorously satisfied at $r = 2m$, so (4.43) together with (4.50) cannot be called a global solution; the "neck" is discontinuous. Yet the closeness to a bottle picture solution is tantalizing. There are many ways of removing some of the restrictive assumptions used here which might avoid the discontinuous neck, and they all appear to complicate the algebraic problem significantly. We hope to return to these problems at a later date.

V. SUMMARY AND CONCLUSION

With the imposition of strong spheristasis on a pair of hyperbolic metrics we have shown the existence of nontrivial solutions to the equations, proposed earlier, of a strong and weak gravity theory by use of equation splitting. Attempts to generalize to weak spheristasis is shown to leave apparently undetermined the additional functions that one required. The equation splitting determines a differential problem where knowledge of GR can be made use of, and an algebraic problem which is handled by an earlier analysis. This algebraic analysis leaves only one possibility open involving the exceptional case, and this yields a solution either inside the radius $r = 2m$ with ${}^0\mu + k^2\,{}^1\mu = 0$ (the exceptional case) holding or a solution outside the radius $r = 2m$ with ${}^0\mu - k^2\,{}^1\mu = 0$ holding; it is, of course, clear that these solutions inside and outside $r = 2m$ cannot be continuations of one another, since their respective existences force different conditions on the parameters and the intersection of these conditions is completely trivial (${}^0\mu = {}^1\mu = 0$), negating all coupling of the two fields.

One may, however, continue the solution for ${}^0g_{\cdot\cdot}$ in the usual way, to the larger manifold which includes the region $r > 2m$ by adopting Eddington–Finkelstein coordinates; in these, however, ${}^1g_{\cdot\cdot}$ does not lose its singularity at $r = 2m$ suggesting that there is no coordinate system in which both ${}^0g_{\cdot\cdot}$ and ${}^1g_{\cdot\cdot}$ are simultaneously and continuously extendible. In this case, one may continue ${}^0g_{\cdot\cdot}$ to $r > 2m$ and define ${}^1g_{\cdot\cdot} = 0$ in this region; by doing this we satisfy the field equations in both regions. Thus, vacuum general relativity becomes modified only inside of the Schwarzschild radius.

With regard to the interaction specified by the two vertex terms of the interaction Lagrangian the two cases ${}^0\mu + k^2\,{}^1\mu = 0$ and ${}^0\mu - k^2\,{}^1\mu = 0$ seem to maximize the weakness of coupling with the constraint ${}^0\mu \neq 0$, ${}^1\mu \neq 0$: In an analogous coupling of scalar fields ${}^0\mu + k^2\,{}^1\mu = 0$ actually eliminates the coupling and ${}^0\mu - k^2\,{}^1\mu = 0$ causes the vertices to coalesce. The form of the equation splitting has already restricted the coupling to be very

weak and this restriction forces one of these two cases. In spite of this extreme weakness of coupling, however, the solution suggests some aspects of the sought *bottle picture* solutions that were described in Refs. 1,2.

We note that if ${}^1\Lambda$ were chosen to be ${}^1\Lambda < 0$, the corresponding pair of solutions would exhibit an infinite number of singular surfaces. Such a solution does not appear to have any relation to the bottle picture and so it is omitted.

ACKNOWLEDGMENTS

The author wishes to express his thanks to Dr. John Cyranski for stimulating discussions, Dr. Stephen A. Fulling for his critical reading of a first draft manuscript and Miss Joyce Miezin for her usual careful and meticulous typing of the paper. I also thank Professor David Lovelock and R.J. Mckellar for reporting to me an error in the manuscript. Conversations with Professors E. Lubkin, N. Papastamatiou, D. Snider, and H. Umezawa have been most helpful.

[1]W.C. Hammel and E. Lubkin, "'Strong' and 'Weak' Gravity: A Class of Generally Covariant Mixing Models of Spin-2 Neutral Fields," University of Wisconsin-Milwaukee Preprint, UWM-4867-73-7. (A revised version in Nuovo Cimento B **26**, 377 (1975). W.C. Hammel, Ph.D. thesis, University of Wisconsin-Milwaukee (1974).
[2]W. C. Hammel, Nuovo Cimento B **25**, 757 (1975); Ph. D. thesis (1974).
[3]S.N. Gupta, Rev. Mod. Phys. **29**, 334 (1957); Phys. Rev. **96**, 1683 (1954).
[4]C.J. Isham, A. Salam, and J. Strathdee, Phys. Rev. D**3**, 867 (1971); R. Delbourgo, A. Salam, and J. Strathdee, Nuovo Cimento **49**, 595 (1967); Phys. Lett. **22**, 680 (1960); P.C. Aichelburg and R. Mansouri, Nuovo Cimento **10**B, 483 (1972); P.C. Aichelburg, Phys. Rev. D**8**, 377 (1973).
[5]For example, J.K. Lawrence and E.T. Toton, Ann. Phys. **72**, 293 (1972); R. Mansouri and H.K. Urbantke, Commun. Math. Phys. **26**, 301 (1972); Helmut K. Urbantke, J. Math. Phys. **15**, 129 (1974).
[6]J. L. Synge and A. Schild, *Tensor Calculus* (University of Toronto Press, Toronto, 1962); E. Schrödinger, *Space-Time Structure* (Cambridge U.P. Cambridge, 1963).
[7]K. Yano, *The Theory of Lie Derivatives and Its Applications* (North-Holland, Amsterdam, 1955).
[8]S.W. Hawking and G.F.R. Ellis, *The Large Scale Structure of Space–Time* (Cambridge U.P., Cambridge, 1973), Appendix B.
[9]See for example, Ya. B. Zeldovich and I.D. Novikov, *Relativistic Astrophysics* (University of Chicago Press, Chicago, 1971).
[10]These are not the most general coordinate transformations which preserve spherical symmetry but are the most general for our purposes.
[11]A.Z. Petrov, *Einstein Spaces* (Pergamon, New York, 1969).
[12]A theorem which we have announced denying the existence of conformally related solutions in the exceptional case is erroneous.
[13]The constant of integration Z_N has to do with the scaling of R and m; if Z_N is fixed and, further, Λ is fixed, then (4.34) determines a linear *discrete* mass spectrum.
[14]A. Komar, "Analytic Continuation and Global Theorems in General Relativity" in *Perspectives in Geometry and Relativity*, edited by B. Hoffmann (Indiana U.P., Bloomington, 1966).
[15]A.S. Eddington, Nature **133**, 192 (1924).
[16]D. Finkelstein, Phys. Rev. **110**, 965 (1958).
[17]M.D. Kruskal, Phys. Rev. **119**, 1743 (1960).

A Clifford algebraic approach to superfields and some consequences

C. N. Ktorides

International Centre for Theoretical Physics, Trieste, Italy
(Received 31 July 1974)

Frames provided by Clifford algebras C_n are considered for the purpose of expanding a field multiplet (containing, possibly, both bosons and fermions). After giving a brief—mainly geometrical—description of Clifford algebras, the main tools of the present scheme are introduced: a scalar product in C_n, a conjugation operation, and a "Lorentz covariant derivative." It is described how these Clifford algebraic tools can be applied in order to obtain free massless Lagrangian expressions for a number of field theoretical models. It is also shown how gauge fields can arise within this scheme. It appears possible that the suggested formalism can lead naturally to spinor field operators as "gauge fields." A specific example which can lead to a two-component "gauge spinor" is discussed. Possible lines of investigation which could solidify this potentially rich approach are suggested.

I. INTRODUCTION

The concept of a superfield has recently been introduced by Salam and Strathdee.[1] Briefly, the idea behind a superfield is that it brings together, within a single expression, different kinds of ordinary field operators. Both the Salam and Strathdee superfield expansion as well as a similar one by Fronsdal[2] constitute well-defined and elegant structures which aim to incorporate the central features of the so-called supergauge transformations. The latter have been proposed by Wess and Zumino[3] in the four-dimensional spacetime and have exhibited early signs of a promising approach.[4]

The aforementioned superfield expansion is achieved in terms of *internal* anticommuting variables θ_α, $\alpha = 1, \ldots, 4$. Explicitly, a scalar (supergauge) superfield has the form

$$\phi(x\,\theta) = \varphi(x) + \bar{\varphi}^{\alpha}\theta_{\alpha} + \tfrac{1}{2}\bar{\varphi}^{[\alpha\beta]}\theta_{\alpha}\theta_{\beta} + \tfrac{1}{6}\bar{\varphi}^{[\alpha\beta\gamma]}\theta_{\alpha}\theta_{\beta}\theta_{\gamma} + \tfrac{1}{24}\bar{\varphi}^{[\alpha\beta\gamma\delta]}\theta_{\alpha}\theta_{\beta}\theta_{\gamma}\theta_{\delta}. \tag{1}$$

In (1) the θ's are objects which behave like (if not identified with) constant Majorana spinors.[5]

We consider, in this paper, possible merits of employing a different framework for the expansion of a superfield, namely Clifford algebras. There are several reasons for considering these algebras. Before enumerating some of them it is important to point out that there is a basic difference between the present approach and that suggested by Eq. (1). Namely, that we are employing a frame from the *outside* in order to expand a superfield. In particular, we do not accept the presence of the (anticommuting) internal variables in addition to the spacetime points. Our major motivations are the following:

(a) It has already been observed[6] that the generators of supergauge transformations form a Clifford algebra. Accordingly, the embodiment of a superfield within a Clifford algebraic structure could make possible the description of supergauge transformations as inner transformations, i.e., through Clifford multiplication.

(b) A Clifford algebra constitutes the widest possible algebraic structure which can be built from a given vector space. Now, the latter provides a framework for expanding what one ordinarily calls a vector (e.g., $\alpha\hat{i} + \beta\hat{j}$ is an $O(2)$ vector with real numbers α, β as components). Since we are interested in employing algebraic entities for superfield expansions, it is natural to seek them in algebraic generalizations of frames we are familiar with.

(c) A Clifford algebra C_n of an n-dimensional vector space V_n has dimension 2^n. Thus, for $n=3$ one would expect to accommodate the Wess—Zumino scalar supermultiplet which has eight field components (two scalar, two pseudoscalar, and one Majorana spinor). Similarly, for $n=4$ one expects to accommodate the 16-component Wess-Zumino vector supermultiplet. On the other hand, we shall suggest that even the $n=1$ and $n=2$ Clifford algebras constitute legitimate frameworks for superfield expansions. Quite possibly algebras with $n>4$ could be considered. *It would, then, follow that the framework of Clifford algebras can be employed for a multitude of field theoretical models both of the ordinary field variety and of the mixed supermultiplet variety.* We hope that this last statement can be clarified after we have presented some specific examples.[7]

The paper is organized as follows. In Sec. II we give a very brief description of Clifford algebras concentrating mostly on geometrical interpretations. Our analysis is based on the approach of Ref. 8 and could be omitted by the reader who is familiar with Clifford algebras. We proceed to introduce, in Sec. III, our concept of a superfield as well as the main Clifford algebraic tools which are to be subsequently used. We indicate some results which can be arrived at within our framework. The emphasis is on the generation of free massless Lagrangian expressions for each specific model under consideration. It should become evident that our aim is to incorporate within the proposed Clifford algebraic approach basic field theoretical models (e.g., complex Hermitian field, Dirac spinor) as well as field multiplets which mix bosons and fermions (we restrict our attention to scalar and spin-$\frac{1}{2}$ fields in this paper). In Sec. IV we suggest, through two specific examples, a manner according to which gauge fields can be introduced within the proposed scheme. What should become evident from

our exposition is the fact that Clifford algebras provide us with extremely rich—and, consequently, hard to handle—structures. Accordingly, we list in Sec. V, in the form of an outlook, several questions whose clarification is essential before any strong foundations for the suggested scheme are established. Finally, we give, in an appendix, an explicit derivation of one of the Lagrangian expressions mentioned in Sec. III by means of our Clifford algebraic formalism.

II. ELEMENTS OF CLIFFORD ALGEBRAS

In this section, we shall give a short description of Clifford algebras.[8] We concentrate heavily on geometrical aspects which, hopefully, give more concrete images of what is involved.

Let V_n be a tangent vector space to some n-dimensional manifold $\mathcal{M}$. Our considerations will be restricted to flat manifolds so that, for all practical purposes, $\mathcal{M}$ and V_n can be identified. The vector space V_n is defined over the reals $\mathbb{R}$. Let $\{e_k\}$, $k=1,\ldots,n$, be an orthonormal basis for V_n with some metric structure (V_n can possibly be pseudo-Euclidean). The Clifford algebra $\mathcal{C}_n$ of V_n can be constructed once we introduce an (associative) Clifford product which we formally denote by placing two (Clifford) elements next to each other. As an algebra $\mathcal{C}_n$ will also have a vector space structure which is specified via a rule of addition and the employment of a set of scalars. The Clifford product is distributive over the addition. We put the elements of the basis $\{e_k\}$ into a one-to-one correspondence with elements of $\mathcal{C}_n$ and denote their images (to be called 1-vectors) by E_k, $k=1,\ldots,n$. The E_k constitute only a part of the basis of $\mathcal{C}_n$ and we shall think of them, geometrically, as representing one-dimensional directions. In strict differential geometric language they correspond to 1-forms over the manifold $\mathcal{M}$. We now form all possible independent antisymmetric products E_iE_j, $E_iE_jE_k$, etc., up to $E_1E_2\cdots E_n$. The resulting quantities are called[9] unit 2-vectors, 3-vectors, ..., n-vector, respectively. Note that, to within a sign, only one unit n-vector can be formed.

We think of an r-vector as an oriented r-dimensional volume determined by the r 1-vectors entering the product. The unit n-vector, on the other hand, exhausts the dimensionality of the space and is not thought of as a directed volume, but, rather, as the unit *pseudoscalar* in $\mathcal{C}_n$. In order to justify such a characterization we must first give the complete description of a Clifford product which goes beyond the antisymmetric part we have already mentioned. The Clifford product of any two elements W, Z of $\mathcal{C}_n$ has a symmetric $(\cdot)$ and an antisymmetric (Λ) part:

$$WZ = W\cdot Z + W\Lambda Z. \tag{2}$$

Now, two 1-vectors, which correspond to linearly independent vectors in V_n, have a vanishing symmetric part in their Clifford product so that, e.g., $E_1E_2 \equiv E_1\Lambda E_2$. On the other hand, products like $(E_1)(E_1E_2)$ *do not* vanish in $\mathcal{C}_n$. In particular, E_k^2 is a number identical to the square norm of e_k in V_n. More generally, the product between an r-vector and an s-vector is a combination of $|r-s|$ and $(r+s)$-vectors[10] (the latter vanishes if $r+s>n$ or if the $r+s$ 1-vectors involved correspond to a linearly dependent set in V_n). By introducing the number 1 as a 0-vector in $\mathcal{C}_n$, we finally conclude that $(E_1E_2\cdots E_n)(E_1E_2\cdots E_n)=(E_{12\ldots n})^2=\pm 1$. We shall concentrate exclusively on Clifford algebras for which $(E_{12\ldots n})^2=-1$; hence, the name unit pseudoscalar for $E_{12\ldots n}$. This means that the (flat) metric tensors $G_{\mu\nu}$, $\mu,\nu=1,\ldots,n$ of the V_n must follow the rule:

$$(G_{11})(G_{22})\cdots(G_{nn})=+1 \quad \text{if } n=(4k+2) \text{ or } (4k+3),\ k=0,1,2,3\ldots,$$

and

$$(G_{11})(G_{22})\cdots(G_{nn})=-1 \quad \text{if } n=4k \text{ or } (4k+1),\ k=0,1,2,3,\ldots.$$

One can easily verify this statement by referring to Ref. 10. According to these specifications, some of the simplest Clifford algebras are the complex numbers (underlying vector space is V_1 with metric -1), the algebra of quarternions $(-1,-1)$, the Pauli algebra $(+1,+1,+1)$, and the Dirac algebra $(+1,-1,-1,-1)$. Also note that the $\mathcal{C}_2$ algebra of V_2 (whose metric structure is $(+1,+1)$) is admissible, as well as the $\mathcal{C}_3$ algebra of V_3 (with the metric structure $(+1,-1,-1)$) and the $\mathcal{C}_4$ algebra of V_4 (with metric $(-1,+1,+1,+1)$).

A basis $\{E_J\}$ for $\mathcal{C}_n$ is now formed from all possible unit r-vectors ($r=0,\ldots,n$). One easily finds that there are 2^n of them so that $\mathcal{C}_n$ is a 2^n-dimensional vector space. Finally, we mention that the subset of $\mathcal{C}_n$ consisting of all even r-vectors constitutes a subalgebra $\mathcal{C}_n^+$ which can be identified with $\mathcal{C}_{n-1}$. This identification can be extended to a mapping between $\mathcal{C}_n$ and $\mathcal{C}_{n-1}$.[11]

III. CLIFFORD SUPERFIELDS

As a vector space $\mathcal{C}_n$ needs, besides the operation of addition, a field of scalars for its complete specification. Consider now an irreducible set $\mathcal{G}$ of ordinary (relativistic) quantum field operators acting on a Hilbert space of states $\mathcal{H}$. We assume that this set is defined in some sense. In particular, the only features of the field operators which are of direct interest to us are their $ISL(2,C)$ transformation properties [we assume that the $ISL(2,C)$ group has an irreducible unitary representation $U(a,A)$ in $\mathcal{H}$] and a multiplication rule between them (defined, possibly, via distributions). *To define a superfield, we regard the irreducible set of field operators as the set of scalars over which our Clifford algebra is defined.* Thus, a superfield in $\mathcal{C}_n$ has the expansion

$$\Phi(x)=\sum_{J=1}^{2^n}\varphi_J(x)E_J \tag{3}$$

where the $\varphi_J(x)$ being to $\mathcal{G}$ (some of them could be zero). We shall call $\Phi(x)$ a pure superfield if all the $\varphi_J(x)$ have the same $ISL(2,C)$ transformation properties (say, they are all scalar fields), whereas $\Phi(x)$ will be called a complex- or mixed-superfield if the set $\{\varphi_J(x)\}$ in (3) contains field operators with different $ISL(2,C)$ properties. We shall restrict our attention exclusively to scalar and spin-$\frac{1}{2}$ Majorana field operators.[12]

Our purpose is to employ the Clifford algebraic

framework in order to extract facts concerning the (basic) field operators entering the superfield expansion. We are not interested, at this point, in a theory of superfields *per se*.

The central concept on which we shall be basing our conclusions is that of a scalar product in C_n. This scalar product is defined to be a Clifford and $ISL(2,C)$ scalar *at the same time*. A scalar product will be viewed by us as a (Clifford) invariant. We shall propose that Lagrangian expressions for the field operators entering a given superfield expansion are to be constructed from such invariants. Furthermore, we shall call any transformation which leaves this scalar product unaffected a *general superfield transformation*.

Before we can specify the manner in which such a scalar product is to be formed, we must introduce the first of the tools we employ in our formalism. An operation which takes the basis $\{E_J\}$ into a new permutted one $P\{E_J\}$ along with several (specified for each case) sign reversals is called a *Clifford conjugation* operation. A *conjugation operation* in C_n (omit "Clifford" in front) is a Clifford conjugation accompanied by the replacement of field operators by their $ISL(2,C)$ conjugates.[13] We denote by $\tilde{\Phi}(x)$ the conjugate of a superfield $\Phi(x)$. The scalar product (χ,Φ) of two superfields χ and Φ is defined by

$$(\chi,\Phi)\equiv[\tilde{\chi}\Phi]_{S\text{-}L} \tag{4}$$

where S stands for the 0-vector component and L for that part of this component which is an $ISL(2,C)$ scalar. The prescription, in other words, is to Clifford multiply Φ with the conjugate of χ and keep only the $ISL(2,C)$-scalar part in front of the Clifford 0-vector in the resulting expression.

The second important tool entering our scheme is a *Lorentz covariant superfield derivative*. Its construction is prompted by the fact that any field operator $\varphi(x)$ is always defined on points x in spacetime. Accordingly, the only kind of derivatives one can form is $\partial\varphi(x)/\partial x^\mu$, $\mu=0,\ldots,3$. These are directional derivatives which alter the $ISL(2,C)$ behavior of $\varphi(x)$ [e.g., if $\varphi(x)$ is a scalar field, then $\partial\varphi(x)/\partial x^\mu$ transforms as the μth component of a vector field over the Minkowski space, $M_{3,1}$]. We are led to project $\partial/\partial x^\mu$ on the μth direction of the tangent space to $M_{3,1}$. Recalling that the Dirac matrices γ_μ can be thought of as providing a 4×4 matrix representation of an orthonormal Minkowskian base, we introduce a general frame by $\rho(\gamma_\mu)$ where $\rho(\gamma_\mu)$ is some convenient $n\times n$ representation of an orthonormal Minkowskian base.[14] The defining relation of the $\rho(\gamma_\mu)$ is

$$\{\rho(\gamma_\mu),\rho(\gamma_\nu)\}=\eta_{\mu\nu} \tag{5}$$

where $\{\ ,\ \}$ denotes an anticommutator and $\eta_{\mu\nu}$ is the Minkowski metric tensor. Projecting $\partial/\partial x^\mu$ on $\rho(\gamma_\mu)$ we form the operator

$$\not\partial\equiv\sum_{\mu=0}^{3}\rho(\gamma_\mu)\frac{\partial}{\partial x^\mu}\,. \tag{6}$$

This differential operator will not alter the $ISL(2,C)$ nature of any field operator on which it is acting.

The Lorentz covariant derivatives $\Box\Phi$ of a superfield Φ replaces the field operators $\varphi_J(x)$ in the expansion of Φ by $\not\partial\varphi_J$. One can then consider Φ and $\Box\Phi$ on an equal footing, i.e., as superfields whose components have identical $ISL(2,C)$ content.

Apart from a spin-statistics theorem, for which we have no proof at present but which we shall see acting in our formalism, we have the necessary tools to proceed with the generation of free massless Lagrangian expressions $\mathcal{L}_f$ for pure superfields. The rule is the following:

$$\begin{aligned}\mathcal{L}_f&\equiv([\alpha+E_{12\ldots n}\beta\Box]\phi,\ [\alpha+E_{12\ldots n}\beta\Box]\Phi)\\&\equiv((\alpha+E_{12\ldots n}\beta\Box)\Phi)^2,\end{aligned} \tag{7}$$

where $(\alpha,\beta)=(1,1)$ for spin-$\frac{1}{2}$ and $(\alpha,\beta)=(0,1)$ for scalar pure superfields. In (7) $E_{12\ldots n}$ is the unit pseudoscalar in C_n. Furthermore, if the conjugation operation is given in a matrix form by

$$\begin{pmatrix}1\\E_1\\\vdots\\E_{12\ldots n}\end{pmatrix}\rightarrow\begin{pmatrix}&C&\\&(2^n\times2^n)&\end{pmatrix}\begin{pmatrix}1\\E_1\\\vdots\\E_{12\ldots n}\end{pmatrix}, \tag{8}$$

then C is diagonal for scalar and of the form

$$\begin{pmatrix}&&&\pm1\\&0&\iddots&\\&\iddots&\iddots&\\\pm1&&&0\end{pmatrix}$$

for spin-$\frac{1}{2}$ (Majorana) field operators entering a pure superfield.

Following the prescriptions given by (7) and (8) one

TABLE I.

C_n (metric of V_n)	Superfield	Conjugate matrix	$\mathcal{L}_f=((\alpha+E_{12\ldots n}\beta\Box)\Phi,)^2$
(1) C_1, (-1)	$\varphi_1(x)+i\varphi_2(x)$	$\begin{pmatrix}1&\\&-1\end{pmatrix}$	
	(complex Hermitian)		$(\partial_\mu\varphi_1)^2+(\partial_\mu\varphi_2)^2$
(2) C_1, (-1)	$\psi_1^M(x)+i\psi_2^M(x)$	$\begin{pmatrix}&-1\\-1&\end{pmatrix}$	$-\bar{\psi}\vec{\not\partial}\psi$,
	(4- or 2-component spinor)		$\psi+\psi_1^M+i\psi_2^M+i\psi_2^M\quad(i=\sqrt{-1})$
(3) C_2, $(+1,+1)$ $(E_3=E_{12})$	$\psi_1^M(x)E_0+\psi_2^M(x)E_1$ $+\psi_3^M(x)E_2+\psi_4^M(x)E_3$	$\begin{pmatrix}&&-1&-1\\&+1&&\\-1&&&\end{pmatrix}$	$-\sum_{j=1}^{2}\bar{\psi}_j\vec{\not\partial}\psi_j$, $\psi_1=\psi_1^M+i\psi_2^M,\quad\psi_2=\psi_3^M+i\psi_4^M$

can arrive at the free massless Lagrangians for well-known pure superfield models. We list the results in Table I. [Extensive treatment of these (pure) superfield models appears in Ref. 11.]

Some comments are in order with respect to the table. First of all, we remark that a detailed calculation for the third example is given in the appendix. In the second example, the $\psi_j^M(x)$, $j=1,2$, can stand for either 2- or 4-component Majorana spinors. In both cases the result will be $\mathcal{L}_f = -\bar{\psi}\overleftrightarrow{\partial}\psi$ where $\psi(x)=\psi_1^M(x)+i\psi^M(x)$ is, now, an arbitrary spinor. We note, in passing, that we distinguish between the i in the second column of the table (a Clifford direction) and the i in the fourth column (conventional $i=\sqrt{-1}$). In connection with the third example we note that the Clifford basis (E_0, E_1, E_2, E_3) corresponds to $E_0=1$, where E_1, E_2 are the 1-vectors and $E_3=E_1E_2$ is the unit pseudoscalar. The complete multiplication table is given in the appendix. Finally, we mention that mass terms in Lagrangian expressions can also be treated in a similar manner, i.e., as scalar products in C_n, but we have omitted them for simplicity. In particular, a mass term for the first example can arise through the Clifford invariant $(m\Phi, m\Phi)$. For the two spinorial examples above a "spin-statistics" stipulation is needed before a similar expression can be adopted for a mass term. A more extensive discussion on mass terms will appear in Ref. 11.

Identifying C_1 with C_2 and extending this identification to a mapping between C_1 and C_2, we note the correspondences $1\to 1$, E_1 and $i\to E_2$, E_3. We can now combine the two C_1 superfields of the preceding table into a C_2 superfield given by

$$\Phi(x)=\varphi_1(x)+\psi_1^M(x)E_1+\psi_2^M(x)E_2+\varphi_2(x)E_3. \tag{9}$$

If we reflect the difference $i^2=-1$ and $E_2^2=+1$ into the conjugation matrix by taking C to be

$$C=\begin{pmatrix} +1 & & & \\ & +1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix}, \tag{10}$$

it becomes straightforward to show that the application of (7) to superfield (9) gives

$$\mathcal{L}_f=(\partial_\mu\varphi_1)^2+(\partial_\mu\varphi_2)^2-\bar{\psi}\overleftrightarrow{\partial}\psi, \quad \psi_1^M+i\psi_2^M. \tag{11}$$

The Wess–Zumino free Lagrangian (for their scalar supermultiplet) can be arrived at, in C_3, by combining the following two C_2 superfield models: the 4-component spinor of the table and a 4-scalar superfield model in C_2 with two of these fields being superfluous. The sense in which scalar fields are called superfluous is that they enter a free Lagrangian expression only through mass-like terms (i.e., of the form $\alpha\varphi^2$). Accordingly, the equations of motion yield them equal to zero. It is interesting to observe that what we have called superfluous scalar fields appear within our formalism as spinor-like objects.[15] Explicitly, they enter (7) with $(\alpha,\beta)=(1,1)$ and their conjugation matrix is of the off-diagonal form. The details of derivations, along these lines, of the Wess and Zumino free Lagrangian, as well as free Lagrangians for the other superfields mentioned, are given elsewhere.[11]

IV. ON GAUGE FIELDS

There is a variety of transformations for each C_n case which leaves the scalar product (4) invariant, phase transformations in particular. Consider, in this connection, the superfield $\Phi(x)=\varphi_1(x)+i\varphi_2(x)$ in C_1. An infinitesimal phase transformation localized in space time

$$\Phi\to\Phi'=\exp[i\epsilon(x)]\Phi(x)=\Phi(x)+i\epsilon(x)\varphi_1(x)-\epsilon(x)\varphi_2(x) \tag{12}$$

implies

$$\partial\Phi\to\partial\Phi'=\partial\Phi+i\epsilon(x)\partial\!\!\!/\varphi_1-\epsilon(x)\partial\!\!\!/\varphi_2+\partial\!\!\!/\epsilon(x)(\varphi_1 i-\varphi_2). \tag{13}$$

In order to have a truly covariant derivative [i.e., one transforming like $\Phi(x)$] under such phase transformations, the last term on the right-hand side must vanish. For this purpose, we make the replacement

$$\partial\!\!\!/\to\partial\!\!\!/-eA\!\!\!/, \tag{14}$$

where

$$A\!\!\!/=\rho(\gamma_\mu)A_\mu, \tag{15}$$

$\rho(\gamma_\mu)$ being a 1×1 representation of an orthonormal Minkowskian frame (recall previous discussion). If we require that A_μ transform according to

$$A\!\!\!/\to A\!\!\!/-\frac{\epsilon(x)^{-1}}{e}\partial\!\!\!/\epsilon(x), \tag{16}$$

then we easily find that the last term in (13) vanishes. Consequently, $\partial\Phi$ transforms like Φ.

A more interesting transformation with spacetime dependent parameters can be introduced in connection with the C_2 model given by Eq. (9). Consider the infinitesimal spinorial transformation[16] of the subspace $\psi_1^M E_1+\psi_2^M E_2$ given by

$$\psi_1^M E_1+\psi_2^M E_2\to\exp(\bar{\alpha}^M(x)E_1)(\psi_1^M E_1+\psi_2^M E_2)=\psi_1^M E_1+\psi_2^M E_2+\alpha^M\psi_1^M+\bar{\alpha}^M\psi_2^M E_3 \tag{17}$$

where ψ_1^M, ψ_2^M, and α^M are 2-component (Majorana) spinors. Then,

$$\Phi(x)\to\Phi(x)'=\Phi(x)+\bar{\alpha}^M\psi_1^M+\bar{\alpha}^M\psi_2^M. \tag{18}$$

On the other hand,

$$\partial\Phi(x)\to\partial\Phi(x)'=\partial\Phi(x)+\bar{\alpha}^M\partial\!\!\!/\psi_1^M+\bar{\alpha}^M\partial\!\!\!/\psi_2^M E_3+\partial\!\!\!/\bar{\alpha}^M(\psi_1^M+\psi_2^M E_3). \tag{19}$$

We *suggest* the following substitution:

$$\partial\!\!\!/\to\partial\!\!\!/-l^2\overline{(\bar{\nu}\nu)} \tag{20}$$

where $\bar{\nu}\nu$ is a scalar Hermitian field formed by the spinor $\nu(x)$ and its conjugate. Furthermore,

$$\overline{(\bar{\nu}\nu)}=(\bar{\nu}\rho(\gamma_\mu)\nu)\rho(\gamma^\mu) \tag{21}$$

with $\rho(\gamma_\mu)$ being 2×2 representations of a Minkowsian orthonormal frame.[17]

Obviously, $\overline{(\bar{\nu}\nu)}$ should transform by

$$\overline{(\bar{\nu}\nu)}\to\overline{(\bar{\nu}\nu)}-\frac{1}{l^2}\partial\!\!\!/\bar{\alpha}^M(x)(\bar{\alpha}^M(x))^{-1} \tag{22}$$

where $(\bar{\alpha}^M)^{-1}$ is actually a column spinor which satisfies

$$\bar{\alpha}^M(\bar{\alpha}^M)^{-1}=1. \tag{23}$$

It easily follows from the substitution (20) that the last term in (19) will vanish and $\not{\partial}\Phi$ will transform as Φ. One can deduce individual transformation properties of the ν's which combine to give (22) (to the 1st order). We intend to look into these matters in more detail elsewhere.

Now, the spinor $\nu(x)$, which we take to be of dimension $-\frac{1}{2}$, can be thought of as a fermion (spin-$\frac{1}{2}$) "gauge field". The substitution (20), when inserted into (11), gives rise to the (interaction) term $l^2\overline{\nu}\rho(\gamma_\mu)\nu\overline{\psi}\rho(\gamma_\mu)\psi$. We suggest that the coupling constant l^2 is *dimensionless* by taking $[\nu]=-\frac{1}{2}$, $[\psi]=-3/2$. One could say that the fermion "gauge field" stands (in dimension) with respect to the observable spinor as the photon field A_μ stands to the observable $F_{\mu\nu}$. This, of course, is not a new idea. It has been pursued in the past by Dürr and Winter[18] along different lines. Actually, these authors have considered *all* spinors entering a theory as having dimension-$\frac{1}{2}$. Here, we distinguish between "gauge spinors" and "matter spinors."

Our example has been restricted to the 2-component spinorial case and is probably not the most general. Hopefully, arguments along parallel lines can lead to similar conclusions in connection with a 4-component "gauge spinor." Added difficulties in this case could arise with respect to chiral projections.

V. CONCLUSIONS

We have described a formalism which seems capable of yielding (Clifford) invariant expressions (Lagrangians) that cover a wide spectrum of field theoretical models. This formalism can be used to describe both pure and mixed superfields. The present scheme also seems to provide a rich body of transformations which leave our scalar products invariant. Finally, it is capable of introducing gauge fields in a simple straightforward manner. Detailed procedures and calculations concerning the aforementioned matters are subjects of forthcoming papers.

On the other hand, it must be pointed out that the Clifford algebraic structure, as we have presented it, remains largely a formalism. Physical background will begin to crystallize only after much more work has been done. We want to suggest some of the wide open problems whose clarification should have a lot to say about the foundations (or lack of them) of our approach.

(a) A complete justification of the various mathematical objects which enter the scheme is of essential importance. In particular, our crude geometrical interpretations need refinement.

(b) Investigations as to whether the Hilbert space of states can be split in a way similar to the one we have used to expand superfields would be extremely useful. If, indeed, this could be done, a much closer relationship than has been implied would exist between the Hilbert space of states and the superfields.

(c) A classification of the general superfield transformations for each given superfield model is of utmost importance. This is a huge task. To appreciate what is involved here, just consider the case of the C_1 superfield given by $\Phi(x)=\varphi_1(x)+i\varphi_2(x)$. In Fig. 1 we place, on the complex plane ($\equiv C_1$), the product $\tilde{\chi}\Phi$ whose real part gives (χ,Φ). Any transformation which takes $\tilde{\chi}\Phi$ somewhere on the line L is a general superfield transformation. A phase transformation, in particular, leaves $\tilde{\chi}\Phi$ where it is.

(d) A formulation and proof of a spin-statistics theorem, which seems to be acting in our scheme, is essential.

(e) The generation of mass and/or interaction Lagrangian terms is, of course, vital. In principle, they should enter through invariants like $(m\Phi, m\Phi)$, $(\sqrt{g}\,\Phi\Phi,\ \sqrt{g}\,\Phi\Phi)$, but it could be that some complications are involved, especially when it comes to mixed superfields.

(f) The Lorentz content of our formalism is to be found in our definition of a scalar product. We can afford to adopt this attitude for the moment. On the other hand, should the enquiries in connection with (b) give positive answers, we would have to think more seriously about the Lorentz transformation properties of Clifford directions. In Ref. 11 we consider this question in more detail.

It is hoped that our suggestion that Clifford algebras can provide a framework for the description of superfields can be taken seriously. Investigations along the lines set by this paper are at present in progress.

ACKNOWLEDGMENTS

The author is especially indebted to Dr. A.J. Phares and Professor N.K. Kofinti for several valuable suggestions. Helpful conversations with Professor C. Fronsdal, Dr. J. Strathdee, and Dr. S.T. Ali are also acknowledged. The author is grateful to Professor Abdus Salam, the International Atomic Energy Agency and UNESCO for hospitality at the International Centre for Theoretical Physics, Trieste.

APPENDIX

We shall show, in this appendix, how our formula (7) leads to the Lagrangian for a pure spinorial superfield in C_2. The superfield expansion is

$$\psi(x)=\psi_1^M(x)E_0+\psi_2^M(x)E_1+\psi_3^M(x)E_2+\psi_4^M(x)E_3, \qquad \text{(A1)}$$

where (E_0,E_1,E_2,E_3) forms a base for C_2 ($E_0=1$ and E_3 is the unit pseudoscalar). The multiplication table is the following:

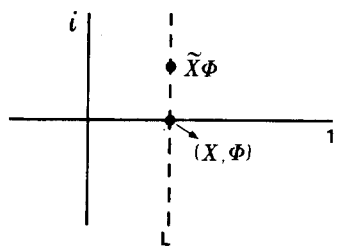


FIG. 1. Illustration of the domain of general superfield transformations in C_1.

	E_0	E_1	E_2	E_3
E_0	E_0	E_1	E_2	E_3
E_1	E_1	E_0	E_3	E_2
E_2	E_2	$-E_3$	E_0	$-E_1$
E_3	E_3	$-E_2$	E_1	$-E_0$

The adjoint of $\psi(x)$ is given, according to our specifications of Sec. III, by

$$\tilde{\Psi} = -\bar{\psi}_1^M(x)E_3 - \bar{\psi}_2^M(x)E_2 + \bar{\psi}_3^M(x)E_1 - \bar{\psi}_4^M(x)E_0. \tag{A2}$$

Finally, we need $\partial\!\!\!/\,\psi$ and $\partial\!\!\!/\,\bar{\psi}$. They are given by

$$\partial\!\!\!/\,\Psi = \partial\!\!\!/\,\psi_1^M E_0 + \partial\!\!\!/\,\psi_2^M E_1 + \partial\!\!\!/\,\psi_3^M E_2 + \partial\!\!\!/\,\psi_4^M E_3 \tag{A3}$$

and

$$\partial\!\!\!/\,\Psi = -\partial\!\!\!/\,\psi_1^M E_3 - \partial\!\!\!/\,\psi_2^M E_2 + \partial\!\!\!/\,\psi_3^M E_1 - \partial\!\!\!/\,\bar{\psi}_4^M E_0. \tag{A4}$$

We recall that the ψ_j^M, $j=1,\ldots,4$, are Majorana spinors Accordingly, we can use the Majorana properties[1]

$$\bar{\psi}^M\psi'^M = \bar{\psi}'^M\psi^M, \quad \partial\!\!\!/\,\bar{\psi}^M\partial\!\!\!/\,\psi'^M = \partial\!\!\!/\,\bar{\psi}'^M\partial\!\!\!/\,\psi^M. \tag{A5}$$

Following formula (7) we must now form

$$\Psi + E_3\partial\!\!\!/\,\Psi = (\psi_1^M - \partial\!\!\!/\,\psi_4^M)E_0 + (\psi_2^M + \partial\!\!\!/\,\psi_3^M)E_1 + (\psi_3^M - \partial\!\!\!/\,\psi_2^M)E_2 + (\psi_4^M + \partial\!\!\!/\,\psi_1^M)E_3 \tag{A6}$$

and

$$\widetilde{(\Psi + E_3\partial\!\!\!/\,\Psi)} = -(\bar{\psi}_1^M - \partial\!\!\!/\,\bar{\psi}_4^M)E_3 - (\bar{\psi}_2^M + \partial\!\!\!/\,\bar{\psi}_3^M)E_2 + (\bar{\psi}_3^M - \partial\!\!\!/\,\bar{\psi}_2^M)E_1 - (\bar{\psi}_4^M + \partial\!\!\!/\,\bar{\psi}_1^M)E_0, \tag{A7}$$

so that

$$\begin{aligned}\mathcal{L}_f &= [\widetilde{(\Psi + E_3\partial\!\!\!/\,\Psi)}(\Psi + E_3\partial\!\!\!/\,\Psi)]_{S\text{-}L} \\ &= -\bar{\psi}_4^M\psi_1^M + \bar{\psi}_4^M\partial\!\!\!/\,\psi_4^M - \partial\!\!\!/\,\bar{\psi}_1^M\psi_1^M - \partial\!\!\!/\,\bar{\psi}_1^M\partial\!\!\!/\,\psi_4^M + \bar{\psi}_3^M\psi_2^M + \bar{\psi}_3^M\partial\!\!\!/\,\psi_3^M \\ &\quad - \partial\!\!\!/\,\bar{\psi}_2^M\psi_2^M - \partial\!\!\!/\,\bar{\psi}_2^M\partial\!\!\!/\,\psi_3^M - \bar{\psi}_2^M\psi_3^M + \bar{\psi}_2^M\partial\!\!\!/\,\psi_2^M - \partial\!\!\!/\,\bar{\psi}_3^M\psi_3^M \\ &\quad + \partial\!\!\!/\,\bar{\psi}_3^M\partial\!\!\!/\,\psi_2^M + \bar{\psi}_1^M\psi_4^M + \bar{\psi}_1^M\partial\!\!\!/\,\psi_1^M - \partial\!\!\!/\,\bar{\psi}_4^M\psi_4^M - \partial\!\!\!/\,\bar{\psi}_4^M\partial\!\!\!/\,\psi_1^M.\end{aligned} \tag{A8}$$

Note that *we have only kept the E_0 part of the Clifford product* $\widetilde{(\Psi + E_3\partial\!\!\!/\,\Psi)}(\Psi + E_3\partial\!\!\!/\,\Psi)$ in accordance with our definition of $\mathcal{L}_f$. Using the Majorana properties (A5) we get cancellations leading to

$$\mathcal{L}_f = -\sum_{j=1}^{4}\bar{\psi}_j\overleftrightarrow{\partial\!\!\!/}\psi_j. \tag{A9}$$

Recall, now, from the conventional Dirac theory that we can take

$$\bar{\psi} = (\psi_1^*\ \psi_2^*\ \psi_3^*\ \psi_4^*)\gamma_0. \tag{A10}$$

Writing

$$\psi_\alpha = (\psi_1 + i\psi_2)_\alpha, \quad \alpha = 1,\ldots,4, \tag{A11}$$

we obtain for the ordinary Dirac Lagrangian (omitting factors in front which are inessential to our argument)

$$\mathcal{L} = \bar{\psi}\overleftrightarrow{\partial\!\!\!/}\psi = (\psi_1 - i\psi_2)_\alpha[\overleftrightarrow{A}]_{\alpha\beta}(\psi_1 + i\psi_2)_\beta, \tag{A12}$$

where

$$[\overleftrightarrow{A}]_{\alpha\beta} = (\gamma_0\gamma_\mu)_{\alpha\beta}\overleftrightarrow{\partial}_\mu. \tag{A13}$$

It is not hard to see that the imaginary parts of (A12) vanish:

$$i[\psi_{1\alpha}\overrightarrow{A}_{\alpha\beta}\psi_{2\beta} - \psi_{1\alpha}\overleftarrow{A}_{\alpha\beta}\psi_{2\beta} - \psi_{2\alpha}\overrightarrow{A}_{\alpha\beta}\psi_{1\beta} + \psi_{2\alpha}\overleftarrow{A}_{\alpha\beta}\psi_{1\beta}] = 0. \tag{A14}$$

Indeed, since $(\gamma_0\gamma_\mu)$ is symmetric and since the ψ's anticommute, we have, e.g.,

$$\partial_\mu\psi_{1\alpha}(\gamma_0\gamma_\mu)_{\alpha\beta}\psi_{2\beta} = -\psi_{2\beta}(\gamma_0\gamma_\mu)_{\beta\alpha}\partial_\mu\psi_{1\alpha}, \tag{A15}$$

which shows that the second and third terms in (A14) cancel. Similar arguments hold for the first and fourth terms. We are finally left with just the real contributions in (A12). Explicitly,

$$\mathcal{L} = \bar{\psi}\overleftrightarrow{\partial\!\!\!/}\psi = \bar{\psi}_1\overrightarrow{\partial\!\!\!/}\psi_1 + \bar{\psi}_2\overrightarrow{\partial\!\!\!/}\psi_2. \tag{A16}$$

Returning to our Clifford algebraic result,

$$\mathcal{L}_f = -\bar{\psi}_1^M\overleftrightarrow{\partial\!\!\!/}\psi_1^M - \bar{\psi}_2^M\overleftrightarrow{\partial\!\!\!/}\psi_2^M - \bar{\psi}_3^M\overleftrightarrow{\partial\!\!\!/}\psi_3^M - \bar{\psi}_4^M\overleftrightarrow{\partial\!\!\!/}\psi_4^M, \tag{A17}$$

we can group together, e.g., ψ_1^M and ψ_2^M, into an arbitrary (i.e., not necessarily Majorana) spinor ψ by

$$\psi = \psi_1^M + i\psi_2^M. \tag{A18}$$

Similarly for $\psi_3^M + i\psi_4^M$:

$$\psi' = \psi_3^M + i\psi_4^M. \tag{A19}$$

We then have as a final answer

$$\mathcal{L}_f = -\bar{\psi}\overleftrightarrow{\partial\!\!\!/}\psi - \bar{\psi}'\overleftrightarrow{\partial\!\!\!/}\psi', \tag{A20}$$

which is the expression quoted in the table of Sec. III.

Finally, we remark that in (A18) and (A19) the i stands for $\sqrt{-1}$ rather than the unit pseudoscalar in $\mathcal{C}_1$.

[1]Abdus Salam and J. Strathdee, Nucl. Phys. B 76, 477 (1974).
[2]C. Fronsdal, ICTP, Trieste, Preprint IC/74/21.
[3]J. Wess and B. Zumino, Nucl. Phys. B 70, 39 (1974).
[4]J. Wess and B. Zumino, Phys. Lett. 49B, 52 (1974); J. Iliopoulos and B. Zumino, Nucl. Phys. B 76, 310 (1974); J. Wess and B. Zumino, *ibid.* 78, 1 (1974); S. Ferrara, J. Wess and B. Zumino, Phys. Lett. 51B, 239 (1974).
[5]At least in the sense that $\bar{\varphi}^\alpha(x)\theta_\alpha$ is a Lorentz scalar when $\varphi_\alpha(x)$ is an ordinary Majorana spinor field.
[6]Abdus Salam and J. Strathdee, Nucl. Phys. B 80, 499 (1974).
[7]One immediate implication we must point out is that the Clifford algebraic approach, from our point of view at least, does not exclusively aim to accommodate supergauges. If this was our sole purpose it is doubtful that Clifford algebras would be more successful than the Salam and Strathdee approach to superfields. In this connection, let us also mention that an attractive feature of expansion (1) that we are, naturally, giving up is the presence of the internal variables θ_α which bear some similarity to twistors. In the latter formalism, however, spacetime points play a secondary role, whereas in (1) they seem to be taken on equal footing with the θ's.
[8]D. Hestenes, *Spacetime Algebra* (Gordon and Breach, New York, 1966).
[9]We stick to the nomenclature of Ref. 8 even though it would be more appropriate to call the corresponding objects r-forms.
[10]The Clifford product between an r-vector $(\nu_1\wedge\nu_2\cdots\wedge\nu_r)$ and an s-vector $(u_1\wedge u_2\cdots\wedge u_s)$ (say, $r<s$) is given by

$$(\nu_1\wedge\nu_2\wedge\cdots\wedge\nu_r)(u_1\wedge u_2\wedge\cdots\wedge u_s) = (\nu_1\wedge u_2\wedge\cdots\wedge\nu_r\wedge u_1\wedge\cdots\wedge u_s) + \sum_{all\,perm}(-1)^e[(\nu_1\wedge\cdots\wedge\nu_s)\cdot(u_{j_1}\wedge\cdots\wedge u_{j_s})]u_{j_{s+1}}\wedge\cdots\wedge u_{j_r}$$

where $\epsilon = \begin{cases} +1 & \text{even perm} \\ -1 & \text{odd perm} \end{cases}$. Finally, $[(\nu_1 \wedge \cdots \wedge \nu_s) \cdot (u_1 \wedge \cdots \wedge u_s)]$ is a number given by

$$\det \begin{vmatrix} \nu_s \cdot u_1 & \cdots & u_s \cdot \nu_s \\ \vdots & & \vdots \\ \nu_1 \cdot u_1 & \cdots & \nu_1 \cdot u_s \end{vmatrix}$$

[11]C. N. Ktorides, ICTP, Trieste, Preprint IC/75/47 (1975), submitted for publication.

[12]We *need* Majorana spinor components in superfield expansions because the property $\bar{\psi}\psi' = \bar{\psi}'\psi$ is essential to our proofs.

[13]By conjugate to a field operator $\varphi_{\alpha\beta\ldots}(x)$ we mean a field operator $\bar{\varphi}^{\alpha\beta\cdots}(x)$ such that $\bar{\varphi}^{\alpha\beta\cdots}(x)\varphi_{\alpha\beta\ldots}(x)$ is an ISL(2, C) Hermitian scalar field.

[14]For examples of such representations, see, E.M. Corson, *Introduction to Tensors, Spinors and Relativistic Wave Equations* (Blackie, London, 1953). In particular, a 2×2 representation is mentioned on p. 163; representations of higher dimensionality (given by Kemmer) are exhibited as well (pp. 181, 182).

[15]This observation strengthens our feeling that a spin-statistics theorem is present in our formalism. In particular, the vanishing of the superfluous scalar fields could be considered as a consequence of the fact that these fields obey the wrong statistics.

[16]It can be shown that transformation (17) leaves (7) invariant, given a certain condition imposed on the α_j^M, $j=1,2$. We assume that the corresponding action on $\tilde{\Phi}(x)$ [using the conjugation matrix given by (10)] is given by

$$\bar{\psi}_1^M E_2 + \bar{\psi}_2^M E_1 \to (-\psi_1^M E_2 + \bar{\psi}_2^M E_1)\exp[\alpha^M(x)E_2].$$

[17]Since we have not been systematic, by any means, in introducing $(\bar{\nu}\nu)$, we make no firm commitments as to the nature of this object, i.e., whether it represents a bound state or not. We intend to investigate the question of fermion "gauge field" more thoroughly in the future. Here, we merely want to present an illustration of the possibility of introducing such an object within the framework of our formalism.

[18]H. P. Dürr and N.J. Winter, Nuovo Cimento A 70, 467 (1970); Nuovo Cimento A 7, 461 (1972).

A non-Lie algebraic framework and its possible merits for symmetry descriptions

C. N. Ktorides

International Centre for Theoretical Physics, Trieste, Italy
(Received 25 November 1974; final revision received 13 February 1975)

A nonassociative algebraic construction is introduced which bears a relation to a Lie algebra L paralleling the relation between an associative enveloping algebra and L. The key ingredient of this algebraic construction is the presence of two parameters which relate it to the enveloping algebra of L. The analog of the Poincaré–Birkhoff–Witt theorem is proved for the new algebra. Possibilities of physical relevance are also considered. It is noted that, if fully developed, the mathematical framework suggested by this new algebra should be non-Lie. Subsequently, a certain scheme resulting from specific considerations connected with this (non-Lie) algebraic structure is found to bear striking resemblance to a recent phenomenological theory proposed for explaining CP violation by the K^0 system. Some relevant speculations are also made in view of certain recent trends of thought in elementary particle physics. Finally, in an appendix, a Gell-Mann–Okubo-like mass formula for the new algebra is derived for an $SU(3)$ octet.

I. INTRODUCTION

The effectiveness of Lie algebras in theoretical physics must be considered, by now, as established. Nevertheless, with the deepening of our understanding of problems like symmetry breaking, discrete space–time symmetry violations, etc., the need for a more effective algebraic structure emerges as a plausible alternative. Of course, if one were to adopt such an attitude one should not lose sight of the vast and numerous successes of Lie algebras (and groups) in connection with elementary particles. In this regard, one may recall the so-called Jordan algebras introduced as far back as 1934 by Jordan, von Neumann, and Wigner[1] defined by the identities

$$ab = ba, \tag{1.1a}$$

$$(a^2b)a = a^2(ba), \tag{1.1b}$$

where the algebraic product has been denoted by simply placing two elements next to each other. (Perhaps a more familiar representation of the Jordan product is $\{a, b\}$.[2]) Despite several attempts, the question of possible physical usefulness, either of Jordan algebras or various generalizations of them, remains open at this time with no clear-cut application surfacing so far. Accordingly, it makes more sense that any alternative algebraic structure put forth for improving the *status quo* should have a Lie content in some way or other.

In this respect, the Santilli algebras,[3] proposed in more recent times, seem to be more promising. The important aspect of the Santilli algebras is Lie admissibility, a concept which goes back to Albert.[4]

Briefly, an algebra A is Lie admissible if its product can be utilized, via a commutator, to introduce a new product satisfying the Lie properties. Throughout this paper we adopt the Lie admissibility in Santilli's sense. It is specified as follows. Given an algebra A (product ab) denote $[a, b, c] \equiv (ab)c - a(bc)$, the so-called associator. Santilli's Lie admissible algebras are defined by the relations

$$[a, b, a] = 0 \tag{1.2a}$$

$$[a, b, c] + [b, c, a] + [c, a, b] = 0. \tag{1.2b}$$

Note that the latter is a generalization of the Jacobi identity.

The first realization of the adopted Lie admissibility axioms has been given by Santilli and Soliani.[5] It can be viewed as a combination of Lie and Jordan algebras with two parameters entering. These parameters express the mixing between the two algebras. Explicitly, consider an associative algebra A whose product we denote by placing two elements next to each other. It is well known that the associative product can be used to define the Lie algebra A_L of A. Thus, A_L has the same elements as A but the product in the former is specified by

$$[a, b] = ab - ba. \tag{1.3}$$

The Santilli–Soliani algebra $A(\lambda, \mu)$ is introduced by considering a different generalization of A. In particular, let $A(\lambda, \mu)$ be the same vector space as A but with the product defined via

$$(ab) = \lambda ab + \mu ba. \tag{1.4}$$

The algebra $A(\lambda, \mu)$ so defined has been called the (λ, μ) mutation of A, mutation algebra for short. Obviously, the word mutation characterizes the flexibility one now has due to the adjustability of parameters λ and μ. Note that (1.4) can be rewritten as

$$(ab) = \frac{\lambda - \mu}{2}(ab - ba) + \frac{\lambda + \mu}{2}(ab + ba), \tag{1.5}$$

which illustrates the mixing of the Lie and Jordan algebra products. One sees right away that algebra $A(\lambda, \mu)$ satisfies an "asymptotic condition" in the sense that it reduces directly into a Lie algebra (i.e., A_L) as $\lambda \to +1$, $\mu \to -1$. It can be shown very easily that $A(\lambda, \mu)$ is Lie admissible. The transition from a Lie algebra $\mathcal{L}$ to the more general (Lie admissible) algebraic structure does not entail, of course, the complete abandonment of the Lie framework. We can, in fact, say quite generally [i.e., irrespective of the particular replacement $A_L \to A(\lambda, \mu)$] that the replacement of $\mathcal{L}$ with any Lie admissible algebra U simply implies the embedding of the Lie into the new framework—or the other way round—with $\mathcal{L}$ being identified with U_L. Another in-

teresting feature of the Santilli algebras is that they admit a general analytic formulation whose bracket coincides with product (1.5). Thus, they might be attractive as a methodological tool for investigating interpolating fields.

Some initial applications of algebras $A(\lambda, \mu)$ have been investigated in problems such as[6]: $SU(3)$ symmetry breaking and mass formulas in $A(\lambda, \mu)$ [with $A_L(\lambda, \mu) \approx SU(3)$], classical dissipative systems, plasma instabilities, and quantum mechanical interpolating fields.

In our construction we start from a Lie algebra $\mathcal{L}$. In particular, our original product *is* the Lie product, i.e., we have no associative algebra to begin with. It is well known that one can construct at least one associative algebra from $\mathcal{L}$, namely the universal enveloping algebra $\mathcal{A}$ of $\mathcal{L}$. This algebra could now be used to define a Lie algebra $\mathcal{A}_L$ or a mutation algebra $\mathcal{A}(\lambda, \mu)$ via relations (1.3) and (1.4), respectively. A common feature of $\mathcal{A}$, $\mathcal{A}_L$, and $\mathcal{A}(\lambda, \mu)$ is that they are identical as vector spaces; only their products are different. We also recall[7] that $\mathcal{L}$ is homomorphic, but not isomorphic, to $\mathcal{A}_L$; i.e., not every element of $\mathcal{A}_L$ corresponds to an element of $\mathcal{L}$. Consequently, it is not $\mathcal{L}$ but $\mathcal{A}_L$ which coincides with $\mathcal{A}(1, -1)$.

Our present (λ, μ) generalization is slightly different from the above and is introduced at the level of the tensor product which plays a vital role in the construction of enveloping algebras. The (enveloping) algebra to be constructed will be called by us the universal enveloping mutation algebra (UEMA) of $\mathcal{L}$. In particular, we shall introduce

$$a \otimes' b = \lambda a \otimes b + \mu b \otimes a, \quad a, b \in \mathcal{L}. \tag{1.6}$$

Now, (1.6) does not actually give the UEMA product itself. In fact, the UEMA of $\mathcal{L}$ [to be denoted by $\mathcal{U}(\lambda, \mu)$] is formed as a quotient space of the tensor algebra whose product is $\otimes'$. The point is that (1.6) helps turn $\mathcal{U}(\lambda, \mu)$ into a Lie admissible algebra in Santilli's sense. Thus, we presume that we have here a new realization of identities (1.2a, b) for a Lie admissible algebra which constitutes a true nonassociative enveloping algebra of a Lie algebra. Conceivably, this property might represent a significant contribution for possible physical applications of algebras defined by the aforementioned Lie admissibility identities. Furthermore, while algebras $A(\lambda, \mu)$ were investigated for the limit

$$A(\lambda, \mu) \underset{\substack{\lambda \to +1 \\ \mu \to -1}}{\to} \mathcal{L},$$

we are now lead to focus our attention on the other significant limit

$$\mathcal{U}(\lambda, \mu) \underset{\substack{\lambda \to +1 \\ \mu \to 0}}{\to} \mathcal{A},$$

where $\mathcal{A}$ is the (associative) enveloping algebra of $\mathcal{L}$.

We shall devote Sec. II to the construction of a UEMA of a Lie algebra as well as the derivation of some of its properties. One central part of this section will be the proof of the analog to the powerful Poincaré–Birkhoff–Witt theorem for a UEMA known, until now, for a universal enveloping algebra of a Lie algebra only.

We should also mention some physical motivations for our work. It is actually hard to be concrete and specific on this point, since our main concern right now are introductory mathematical properties of the UEMA. One *could* speculate, on the other hand, as to the kind of uses one can make of two free parameters one has in possession. For example, if one was to give up the idea of Lie groups as forming a certain nucleus for physical descriptions, the possibility opens that a "Lie" asymmetry is some other (non-Lie) group's (or algebra's) symmetry. In this sense, the $\mathcal{U}(\lambda, \mu)$ could be used, e.g., to write invariant [from the $\mathcal{U}(\lambda, \mu)$ point of view] Lagrangian expressions which exhibit Lie asymmetrical terms parametrized through λ and μ. It must be admitted, however, that such an alternative is far from obvious at this point since one lacks the parallel of the Lie group that goes with a UEMA. After all, it is the representations of the group which really matter in constructing invariant Lagrangian expressions. We shall deal with such problems in Sec. III. In particular, we shall exhibit a possible connection between a certain $\mathcal{U}(\lambda, \mu)$-invariance scheme and a superweak phenomenological Lagrangian formulation suggested recently by Hsu[8] to account for CP violation by the K^0 system.

It is needless to emphasise that our understanding of questions revolving around physical applications is still rudimentary. Much more work is necessary before one can talk with confidence about the importance of Lie algebraic generalizations of the sort we are presently introducing. Some speculations along lines concerning relevance to physics are made in Sec. IV.

In the appendix we present another possible application of a UEMA along more orthodox lines. By restricting $\mathcal{L}$ to $SU(3)$ we subsequently derive, within the framework of its UEMA, a Gell-Mann–Okubo-like mass formula which contains the parameters λ and μ in a way similar, but not equivalent, to the previous derivation in the framework of algebras $A(\lambda, \mu)$.[6]

II. UNIVERSAL ENVELOPING MUTATION ALGEBRA OF A LIE ALGEBRA

A. Construction and elementary properties

Throughout this section and the rest of the paper, the term "algebra" stands for associative algebra with an identity element. At times, however, we shall use the adjective "associative" for emphasis. On the other hand, a nonassociative algebra should read as "not necessarily associative." Obviously, we are committed to use the adjective "nonassociative" whenever we are referring to a not necessarily associative algebra which lacks any other kind of characterization. A similar comment holds for the "Lie" characterization of an algebra—whenever it applies.

Suppose we are given an algebra $\mathcal{A}$ whose product is formally denoted by placing two elements of $\mathcal{A}$ next to each other. We can always form the Lie algebra $\mathcal{A}_L$ of $\mathcal{A}$ by introducing the Lie product $[a, b] \equiv ab - ba$, $a, b \in \mathcal{A}$ or $\mathcal{A}_L$. From now on the Lie algebra of an (associative) algebra will be denoted by the subscript L.

Let us recall the definition of the universal enveloping algebra $\mathcal{A}$ of a Lie algebra $\mathcal{L}$ through its construction.[9]

An abstract definition can be given as well; however, it holds true that any two universal enveloping algebras of $\mathcal{L}$ must be isomorphic.[10] Consequently, the definition by construction is perfectly general. To construct the universal (associative) enveloping algebra for a Lie algebra $\mathcal{L}$, one first forms the (associative) tensor algebra of $\mathcal{L}$,

$$\mathcal{T}(\mathcal{L})=F\oplus\mathcal{L}\oplus\mathcal{L}\otimes\mathcal{L}\oplus\cdots, \tag{2.1}$$

where F is the field of scalars over which $\mathcal{L}$ is defined as a vector space. Next, one considers the ideal $\mathcal{R}$ spanned by all elements of the form

$$[l_1l_2]-l_1\otimes l_2+l_2\otimes l_1, \quad l_1,l_2\in\mathcal{L}, \tag{2.2}$$

and $[l_1l_2]$ denotes the Lie product. Finally, the algebra $\mathcal{A}$ formed by the quotient $\mathcal{A}\equiv\mathcal{T}/\mathcal{R}$ can be shown (or, alternatively, can be defined) to be the (associative) universal enveloping algebra of $\mathcal{L}$.

Suppose now that we form a different kind of a tensor algebra $\mathcal{T}'(\mathcal{L})$, which is not even associative, by the replacement $\otimes\rightarrow\otimes'$ where

$$l_1\otimes' l_2=\lambda l_1\otimes l_2+\mu l_2\otimes l_1, \quad \lambda,\mu\in F. \tag{2.3}$$

Explicitly,

$$\mathcal{T}'(\mathcal{L})=F\oplus\mathcal{L}\oplus\mathcal{L}\otimes'\mathcal{L}\oplus\cdots. \tag{2.4}$$

As mentioned before, $\mathcal{T}'(\mathcal{L})$ is not associative. One can easily verify that fact by observing, e.g., that $(l_1\otimes' l_2)\otimes' l_3$ is not necessarily identical with $l_1\otimes'(l_2\otimes' l_3)$, $l_j\in\mathcal{L}$, $j=1,2,3$. Furthermore, it follows from (2.3) that $\mathcal{T}'(\mathcal{L})$ is a subspace of $\mathcal{T}(\mathcal{L})$.

Consider now the subspace $\mathcal{R}'$ of $\mathcal{T}'(\mathcal{L})$ which is spanned by all elements of the form

$$[l_1l_2]-l_1\otimes' l_2+l_2\otimes' l_1, \quad l_1,\ l_2\in\mathcal{L}. \tag{2.5}$$

$\mathcal{R}'$ is an ideal in $\mathcal{T}'(\mathcal{L})$ under the $\otimes'$ product. Indeed, consider the mapping $\mathcal{T}(\mathcal{L})\rightarrow\mathcal{T}'(\mathcal{L})$ according to which $\mathcal{L}\rightarrow\mathcal{L}$ and $\otimes\rightarrow\otimes'$. Then $\mathcal{R}\rightarrow\mathcal{R}'$ and $A\otimes\mathcal{R}\rightarrow A'\otimes'\mathcal{R}'$ where the difference between A and A' is the following. If $A=a_1\otimes\cdots\otimes a_n$, $a_j\in\mathcal{L}$, $j=1,\ldots,n$, then $A'=a_1\otimes'\cdots\otimes' a_n$.[11] But since $\mathcal{R}$ is an ideal in $\mathcal{T}(\mathcal{L})$, then $A\otimes\mathcal{R}\subseteq\mathcal{R}$. It follows that $A'\otimes'\mathcal{R}'\subseteq\mathcal{R}'$ and, therefore [since for every A' in $\mathcal{T}'(\mathcal{L})$ there exists its pre-image A in $\mathcal{T}(\mathcal{L})$], $\mathcal{R}'$ is an ideal of $\mathcal{T}'(\mathcal{L})$.

We can now form the quotient

$$\mathcal{U}(\lambda,\mu)\equiv\mathcal{T}'(\mathcal{L})/\mathcal{R}'. \tag{2.6}$$

$\mathcal{U}(\lambda,\mu)$ is a nonassociative algebra. Indeed, let $A,\ B,\ C\in\mathcal{U}(\lambda,\mu)$. Explicitly, $A=a+\mathcal{R}'$, $B=b+\mathcal{R}'$, $C=c+\mathcal{R}'$, $a,b,c\in\mathcal{T}'(\mathcal{L})$. Denote the product in $\mathcal{U}(\lambda,\mu)$ by $\times$. Now,

$$(A\times B)\times C=(a\otimes' b)\otimes' c+\mathcal{R}'=\lambda^2 a\otimes b\otimes c+\mu\lambda(b\otimes a\otimes c + c\otimes a\otimes b)+\mu^2 c\otimes b\otimes a+\mathcal{R}', \tag{2.7}$$

whereas

$$A\times(B\times C)=\lambda^2 a\otimes b\otimes c+\mu\lambda(a\otimes c\otimes b+b\otimes c\otimes a) + \mu^2 c\otimes b\otimes a+\mathcal{R}'. \tag{2.8}$$

The fact that $\mathcal{U}(\lambda,\mu)$ is nonassociative means that when we write, e.g., $A\times B\times C$ we do not have a uniquely defined element. Since both $(A\times B)\times C$ and $A\times(B\times C)$ belong to $\mathcal{U}(\lambda,\mu)$, the expression $A\times B\times C$ will stand for either one of the above two elements whenever no confusion can arise. Later on, however, we shall find it necessary to be specific as to how an expression like, e.g., $A\times B\times C\times D\cdots$ is meant to be organized. It also follows from its definition that $\mathcal{U}(\lambda,\mu)$ is a subset of $\mathcal{A}$; we can think of $\mathcal{U}(\lambda,\mu)$ as the restriction of $\mathcal{A}$ to a hyperplane parametrized by λ and μ.

We next prove the following.

Lemma 2.1: $\mathcal{U}(\lambda,\mu)$ is Lie admissible.

Proof: We form $\mathcal{U}_L(\lambda,\mu)$ from $\mathcal{U}(\lambda,\mu)$ as follows: The elements of $\mathcal{U}_L(\lambda,\mu)$ are the same as those of $\mathcal{U}(\lambda,\mu)$. The operation in $\mathcal{U}_L(\lambda,\mu)$ is denoted by $[A,B]$ and is specified by $[A,B]=A\times B-B\times A$.

It is evident that $\mathcal{U}_L(\lambda,\mu)$ is a vector space over F with the Lie product $[\ ,\ ]$ being distributive over addition and $\alpha[A,B]=[(\alpha A),B]=[A,(\alpha B)]$, $\alpha\in F$, so that, to begin with, $\mathcal{U}_L(\lambda,\mu)$ is an algebra.

Now, $A=a+\mathcal{R}'$, $B=b+\mathcal{R}'$, $a,b\in\mathcal{T}'(\mathcal{L})$. It follows that

$$A\times B=a\otimes' b+\mathcal{R}'=\lambda\, a\otimes b+\mu\, b\otimes a+\mathcal{R}' \tag{2.9}$$

and

$$B\times A=\lambda\, b\otimes a+\mu\, a\otimes b+\mathcal{R}'. \tag{2.10}$$

So that

$$[A,B]=(\lambda-\mu)(a\otimes b-b\otimes a)+\mathcal{R}'. \tag{2.11}$$

One can now verify the Jacobi identity explicitly or recognize from (2.12) that one basically has a Lie product definition arising from an associative algebra. [Recall that $\mathcal{T}(\mathcal{L})$, with multiplication $\otimes$, is associative.] It is also trivially evident that $[A,A]=0$ since $\mathcal{R}'$ is the zero element of $\mathcal{U}_L(\lambda,\mu)$.

Note that if $l_1,\ l_2\in\mathcal{L}$, then

$$[L_1,L_2]=(\lambda-\mu)(l_1\otimes l_2-l_2\otimes l_1)+\mathcal{R}' \tag{2.12}$$

is the corresponding product in $\mathcal{U}_L(\lambda,\mu)$. There exists, now, a homomorphism between $\mathcal{L}$ and $\mathcal{U}_L(\lambda,\mu)$ according to which $[l_1l_2]$ goes over to $[L_1,L_2]$. Indeed, it is no different a fact to establish than the homomorphism between $\mathcal{L}$ and the Lie algebra $\mathcal{A}_L$ of the universal enveloping algebra of $\mathcal{L}$. The latter is a homomorphism which takes $[l_1l_2]$ to $l_1\otimes l_2-l_2\otimes l_1+\mathcal{R}$, where $\mathcal{R}$, we recall, is an ideal in $\mathcal{T}(\mathcal{L})$. Similar arguments would hold for (2.12). In particular, it is very simple to show that, in our case of interest, the Jacobi identity $[[l_1l_2]l_3]+[[l_2l_3]l_1]+[[l_3l_1]l_2]=0$ goes into the zero element (i.e. $\mathcal{R}'$) of $\mathcal{U}_L(\lambda,\mu)$.

The Lie admissible algebra $\mathcal{U}(\lambda,\mu)$ so constructed will be called by us a universal enveloping mutation algebra (UEMA) for $\mathcal{L}$. The reasoning behind the word "mutation" is that one can think of $\mathcal{U}(\lambda,\mu)$ as changing according to the parameters λ and μ. In particular, if $\mathcal{A}$ is the universal enveloping algebra for $\mathcal{L}$, it follows that

$$\mathcal{U}(\lambda,\mu),\ \mathcal{U}_L(\lambda,\mu)\underset{\lambda\rightarrow1,\,\mu\rightarrow0}{\rightarrow}\mathcal{A},\mathcal{A}_L. \tag{2.13}$$

B. The analog of the Poincare-Birkhoff–Witt theorem for a UEMA

The Poincarê–Birkhoff–Witt theorem (PBW for brevity) is of great significance in connection with universal enveloping algebras. From the point of view of physics, it constitutes the underlying mathematical result on which the derivation of the well-known Gell-Mann–Okubo mass formula rests.

The motivations behind this theorem are related to the desire to order the tensor product expansions of $T(L)$. For example, let $e_1, \ldots, e_n$ be an ordered basis for L and consider $L \otimes L$. Of course, both $e_1 \otimes e_2$ and $e_2 \otimes e_1$ belong to $L \otimes L$. On the other hand, when one views $L \otimes L$ as part of the tensor algebra $T(L)$, one wants to have available a basis for $L \otimes L$ possessing a certain order. The desired ordering hinges on the way in which the e_j's enter the basis elements $e_j \otimes e_k$ $(j, k = 1, \ldots, n)$ of $L \otimes L$. Suppose, in particular, that we want the indices to increase on the right only, i.e., the basis for $L \otimes L$ to be constituted by elements of the form $e_j \otimes e_k$ with $j \leq k$ exclusively. One must ascertain that the elements of the form $e_j \otimes e_k$ with $k < j$ are also covered by such a basis.

We call *standard* any monomial $e_{j_1} \otimes \cdots \otimes e_{j_k}$ (degree k) for which the indices strictly appear in an increasing order from right to left, i.e., $j_s \leq j_r$ if $s \leq r$. Let us note that

$$e_{j_s} \otimes e_{j_{s+1}} - e_{j_{s+1}} \otimes e_{j_s} = [e_{j_s} e_{j_{s+1}}] \bmod R. \tag{2.14}$$

It follows that

$$e_{j_1} \otimes \cdots \otimes e_{j_k} = e_{j_1} \otimes \cdots \otimes e_{j_{s+1}} \otimes e_{j_s} \otimes \cdots \otimes e_{j_k} + e_{j_1} \otimes \cdots \otimes [e_{j_s} e_{j_{s+1}}] \otimes \cdots \otimes e_{j_k} \bmod R. \tag{2.15}$$

In other words, a given monomial can be expressed as a sum of a monomial of the same degree with two members of the product exchanged and a monomial of 1 degree less (mod R). This procedure can be applied successively in order to express a monomial with index $i > 0$ in terms of a standard monomial of the same degree plus a number of monomials each one of 1 degree less. The latter monomials, if not standard, can again be brought into standard form plus monomials of 2 degrees less than that of the original. In conclusion, every element of $T(L)$ can be expressed as an F-linear combination of 1 and standard monomials (mod R).

Finally, the existence of a mapping[12] $\sigma: T(L) \to A$ such that $\sigma(1) = 1$, $\sigma(e_{j_1} \otimes \cdots \otimes e_{j_k}) = E_{j_1} E_{j_2} \cdots E_{j_k}$ if $j_1 \leq j_2 \leq \cdots \leq j_k$ and

$$\sigma(e_{j_1} \otimes \cdots \otimes e_{j_k} - e_{j_1} \otimes \cdots \otimes e_{j_{s+1}} \otimes e_{j_s} \otimes \cdots \otimes e_{j_k}) = \sigma(e_{j_1} \otimes \cdots \otimes [e_{j_s} e_{j_{s+1}}] \otimes \cdots \otimes e_{j_k})$$

opens the way to the PBW theorem. It asserts that the *cosets of* 1 *and of standard monomials form a basis for* $A \equiv T(L)/R$.

This powerful theorem cannot be generalized to $U(\lambda, \mu)$ in a straightforward manner because the latter is nonassociative. We are no longer in a position to liberally make exchanges of two members of a $U(\lambda, \mu)$ monomial. For example, if $j_s > j_{s+1}$, the monomial

$$(\cdots(((\cdots((e_{j_1} \otimes' e_{j_2}) \otimes' e_{j_3}) \cdots) \otimes' e_{j_s}) \otimes' e_{j_{s+1}}) \cdots) \otimes' e_{j_k} \tag{2.16}$$

cannot be expressed, in a manner analogous to (2.15), as

$$(\cdots(((\cdots((e_{j_1} \otimes' e_{j_2}) \otimes' e_{j_3}) \cdots) \otimes' e_{j_{s+1}}) \otimes' e_{j_s}) \cdots) \otimes' e_{j_k} + (\cdots((\cdots((e_{j_1} \otimes' e_{j_2}) \otimes' e_{j_3}) \cdots) \otimes' [e_{j_s} e_{j_{s+1}}]) \cdots) \otimes' e_{j_k} \bmod R'. \tag{2.17}$$

Before we proceed any further, we must decide what should constitute a standard monomial with respect to the operation $\otimes'$ [i.e., in $T'(L)$]. Since we have no *a priori* clues, let us be quite general and declare any monomial in $T'(L)$ standard if the ordering $j_1 \leq j_2 \leq \cdots \leq j_k$ is respected, no matter how the association is made. For example, both $(e_{j_1} \otimes' e_{j_2}) \otimes' e_{j_3}$ and $e_{j_1} \otimes' (e_{j_2} \otimes' e_{j_3})$ are standard with respect to $\otimes'$, provided $j_1 \leq j_2 \leq j_3$. Let us subsequently drop the qualification "with respect to $\otimes'$" for simplicity, unless it is needed for clarification purposes.

Getting back to (2.16) now, we realize that the exchange between e_{j_s} and $e_{j_{s+1}}$ could have taken place along the lines of (2.15) if, instead, we had

$$e_{j_1} \otimes' \cdots \otimes' (e_{j_s} \otimes' e_{j_{s+1}}) \otimes' \cdots \otimes' e_{j_k} \tag{2.18}$$

where we omitted all other brackets (associations) in order to stress the fact that what is important in exchanging e_{j_s} and $e_{j_{s+1}}$ is their being directly multiplied by each other. Accordingly, we are naturally led to the following question: What is the difference between $((a' \otimes' e_{j_s}) \otimes' e_{j_{s+1}}) \otimes' b'$ and $(a' \otimes' (e_{j_s} \otimes' e_{j_{s+1}})) \otimes' b'$ where a', $b' \in T'(L)$? We deal with this question in proving the following.

Lemma 2.2: The difference between the two k-degree monomials in $T'(L)$ $((a' \otimes' e_{j_s}) \otimes' e_{j_{s+1}}) \otimes' b'$ and $(a' \otimes' (e_{j_s} \otimes' e_{j_{s+1}})) \otimes' b'$ (the notation already explained) is a sum of $(k-2)$-degree monomials in $T(L) \bmod R$.

Proof: We have previously evaluated

$$N_1 \equiv ((a' \otimes' e_{j_s}) \otimes' e_{j_{s+1}}) \otimes' b' = [\lambda^2 a' \otimes e_{j_s} \otimes e_{j_{s+1}} + \lambda\mu(e_{j_{s+1}} \otimes a' \otimes e_{j_s} + e_{j_s} \otimes a' \otimes e_{j_{s+1}}) + \mu^2 e_{j_{s+1}} \otimes e_{j_s} \otimes a'] \otimes' b' \tag{2.19}$$

and

$$N_2 \equiv (a' \otimes' (e_{j_s} \otimes' e_{j_{s+1}})) \otimes' b' = [\lambda^2 a' \otimes e_{j_s} \otimes e_{j_{s+1}} + \lambda\mu(a' \otimes e_{j_{s+1}} \otimes e_{j_s} + e_{j_s} \otimes e_{j_{s+1}} \otimes a') + \mu^2 e_{j_{s+1}} \otimes e_{j_s} \otimes a'] \otimes' b'. \tag{2.20}$$

Therefore,

$$N_1 - N_2 = \lambda\mu[e_{j_{s+1}} \otimes a' \otimes e_{j_s} + e_{j_s} \otimes a' \otimes e_{j_{s+1}} - a' \otimes e_{j_{s+1}} \otimes e_{j_s}$$

$$- e_{j_s} \otimes e_{j_{s+1}} \otimes a'] \otimes' b' \tag{2.21}$$

or

$$N_1 - N_2 = \lambda\mu[(e_{j_{s+1}} \otimes a' - a' \otimes e_{j_{s+1}}) \otimes e_{j_s}$$

$$- e_{j_s} \otimes (e_{j_{s+1}} \otimes a' - a' \otimes e_{j_{s+1}})] \otimes' b'. \tag{2.22}$$

Consider the term $e_{j_{s+1}} \otimes a' - a' \otimes e_{j_{s+1}}$. Now, a' has an expansion in $T(L)$ since the product $\otimes'$ is related to $\otimes$. We can now move $e_{j_{s+1}}$ through a' to the right as already suggested by (2.15). Each time we move $e_{j_{s+1}}$ one position to the right we obtain, in addition, a monomial of 1 degree less than the degree of the $T(L)$ monomials in the expansion of a' (mod$R$). When finally, $e_{j_{s+1}}$ is brought all the way to the right, it will cancel with $-a' \otimes e_{j_{s+1}}$, and all that will remain will be a sum of $T(L)$ monomials (mod$R$) of degree $t-1$, where t is the degree of a'. Denoting this sum by the generic symbol c (degree of $c = t-1$), we now have

$$N_1 - N_2 = \lambda\mu[c \otimes e_{j_s} - e_{j_s} \otimes c] \otimes' b'. \tag{2.23}$$

Repeating the same process with $e_{j_s} \otimes c$, we end up (mod$R$) with a sum of a series of monomials of degree $t-2$. Denoting this sum by d we have

$$N_1 - N_2 = d \otimes' b' \bmod R. \tag{2.24}$$

Now b' also has an expansion in $T(L)$ in terms of monomials of degree $k-t$. Consequently, $N_1 - N_2$ is a sum of (not necessarily standard) monomials of degree $k-2$ in $T(L)$ (mod$R$). ∎

Similar arguments hold for the difference $a' \otimes' (e_{j_s} \otimes' (e_{j_{s+1}} \otimes' b')) - a' \otimes' ((e_{j_s} \otimes' e_{j_{s+1}}) \otimes' b')$.

When we bring a $T'(L)$ monomial into the desirable form, we can then apply an argument analogous to that of (2.15). Consider, again, the k degree monomial $((a' \otimes' e_{j_s}) \otimes' e_{j_{s+1}}) \otimes' b'$ for which $j_s > j_{s+1}$ is true. At the expense of introducing a number of $k-2$ – degree $T(L)$ monomials mod$R$, we can work with $(a' \otimes' (e_{j_s} \otimes' e_{j_{s+1}})) \otimes' b'$. It follows that

$$(a' \otimes' (e_{j_s} \otimes' e_{j_{s+1}})) \otimes' b' = (a' \otimes' (e_{j_{s+1}} \otimes' e_{j_s})) \otimes' b'$$

$$+ (a' \otimes' [e_{j_s} e_{j_{s+1}}]) \otimes' b' \bmod R'. \tag{2.25}$$

Finally, the first term on the right-hand side can be brought back to the original format $((a' \otimes' e_{j_{s+1}}) \otimes' e_{j_s}) \otimes' b'$ at the expense of introducing additional $(k-2)$-degree monomials in $T(L)$ mod$R$. The above arguments can be formalized into the following.

Lemma 2.3: A k-degree monomial in $T'(L)$ can be expressed (modR') as an F-linear combination of 1 and standard $T'(L)$ monomials of degree $\leq k$ plus an F-linear combination of 1 and standard $T(L)$ monomials (mod$R$) of degree $\leq (k-2)$.

We have, therefore, found a way to associate together any two members, of a given $T'(L)$ monomial, we wish to exchange. This is the crucial result necessary for our present nonassociative algebraic entity before the analog of the PBW theorem can be extended to $U(\lambda, \mu)$. We are, accordingly, in a position to follow the steps taken for the proof of the theorem in the case of the enveloping algebra. Unfortunately, we are forced to carry, in addition to monomials in $T'(L)$ modR', monomials in $T(L)$ mod$R$ [of a maximum degree smaller by 2 than the corresponding maximum degree of the $T'(L)$ monomials].

Consider now, an arbitrary element $A' = a' + R'$ of $U(\lambda, \mu)$ where $a' \in T'(L)$. We have seen that a' has a decomposition of the form

$$a' = (b'_k + R') + (c_{k-2} + R), \tag{2.26}$$

where b'_k is a sum of $T'(L)$ standard monomials with maximal degree k and c_{k-2} is a sum of $T(L)$ standard monomials of maximal degree $k-2$. We are, thereby, led naturally to our main result which constitutes a generalization of the PBW theorem to UEMA.

Theorem 2.1: An arbitrary element of $U(\lambda, \mu) \equiv T'(L)/R'$ has an F-linear expansion in terms of R' cosets of 1 and standard $T'(L)$ monomials plus R cosets of 1 and standard $T(L)$ monomials. The maximal degree of the latter is 2 less than the maximal degree of the former (monomials).

This theorem is not, obviously, as strong as the corresponding one for universal enveloping algebras. In particular, for a k-degree $U(\lambda, \mu)$ monomial [multiplication in $U(\lambda, \mu)$ denoted by $\times$] the theorem asserts

$$A'_{j_1} \times A'_{j_2} \times \cdots \times A'_{j_k}$$

$$= \sum_{l=k}^{0} \alpha_l A'_{j_{\sigma_1}} \times \cdots \times A'_{j_{\sigma_l}} + \sum_{m=k-2}^{0} \beta_m A_{j_{\rho_1}} A_{j_{\rho_2}} \cdots A_{j_{\rho_m}}, \tag{2.27}$$

where $j_{\sigma_1} \leq j_{\sigma_2} \leq \cdots \leq j_{j_{\sigma_l}}$ and $j_{\rho_1} \leq j_{\rho_2} \leq \cdots \leq j_{\rho_m}$. In (2.27) AB denotes multiplication in $\mathcal{A} = T(L)/R$ $(A, B \in \mathcal{A})$ and $\alpha_l, \beta_m \in F$. Finally, we mention that the particular manner in which the association is made in $A'_{j_1} \times \cdots \times A'_{j_k}$ is irrelevant, given our definition of a $T'(L)$ standard monomial. It is possible, however, to sharpen our definition in view of the fact that any kind of association in the above monomial can be brought, e.g., to the form $(\cdots((A'_{j_1} \times A'_{j_2}) \times A'_{j_3}) \times \cdots) \times A'_{j_k}$ plus a number of $T(L)$ monomials of degree $\leq (k-2)$ mod$R$.

The weakness of the above theorem may be disturbing at first sight. On the other hand, it is possible that Theorem 2.1 could endow one with enough freedom to proceed towards finding "unconventional" physical applications. Thus, the fact that an arbitrary element of our nonassociative algebra $U(\lambda, \mu)$ can be put in a form which contains a standard nonassociative and a standard associative part may prove of relevance, especially when one considers representations of $U(\lambda, \mu)$.

For the applications we have in mind it becomes necessary to possess an associative algebra $\mathcal{A}_{\lambda\mu}$ into which one can map elements of $U(\lambda, \mu)$. This becomes particularly important when one considers representations of $U(\lambda, \mu)$ which act as transformations on a vector space. Hopefully, this point will be elucidated by the

examples in Sec. III as well as that in the appendix. In particular, given the $U(\lambda,\mu)$ product $A\times B$ (A and B may themselves be monomials with more than one factor), we want to be in a position to write

$$A\times B=\lambda A'B'+\mu B'A', \tag{2.28}$$

where A', B' are the images of A, B in $\mathcal{A}_{\lambda\mu}$ and where we have denoted the (associative) product in $\mathcal{A}_{\lambda\mu}$ by placing two of its elements next to each other. As we shall argue later, $\mathcal{A}_{\lambda\mu}$ can be viewed as similar to the universal enveloping algebra $\mathcal{A}$ of $\mathcal{L}$. More generally, $\mathcal{A}_{\lambda\mu}$ can be thought of as an (associative) algebra of $\mathcal{L}$ monomials. We shall have further comments to make about $\mathcal{A}_{\lambda\mu}$ later on.

The feasibility of expression (2.28) can be shown by the following:

Theorem 2.2: There exists a mapping $\sigma: U(\lambda,\mu)\to\mathcal{A}_{\lambda\mu}$, where $\mathcal{A}_{\lambda\mu}$ is an associative algebra of $\mathcal{L}$ monomials, such that

$$\sigma(A\times B)=\lambda\sigma(A)\,\sigma(B)+\mu\sigma(B)\,\sigma(A),\quad A,B\in U(\lambda,\mu). \tag{2.29}$$

Proof: Consider the mapping $\rho: T(\mathcal{L})\to\mathcal{A}_{\lambda\mu}$ such that

$$\rho(a\otimes b)=\rho(a)\rho(b)/(\lambda-\mu). \tag{2.30}$$

It follows that

$$\begin{aligned}&\rho([l_1l_2]-l_1\otimes' l_2+l_2\otimes' l_1)\\&\quad=\rho\{[l_1l_2]-(\lambda-\mu)l_1\otimes l_2+(\lambda-\mu)l_2\otimes l_1\}\\&\quad=\rho([l_1l_2])-\rho(l_1)\rho(l_2)+\rho(l_2)\rho(l_1).\end{aligned} \tag{2.31}$$

As we shall remark in more detail in the next section, whenever we are restricting our considerations on the Lie algebra of a Lie group we can always write

$$\rho([l_1l_2])=\rho(l_1)\rho(l_2)-\rho(l_2)\rho(l_1). \tag{2.32}$$

It then follows that

$$\rho\{[l_1l_2]-l_1\otimes' l_2+l_2\otimes' l_1\}=0. \tag{2.33}$$

Consequently,

$$\rho(A\times B)=\frac{\lambda}{\lambda-\mu}\,\rho(a)\rho(b)+\frac{\mu}{\lambda-\mu}\,\rho(b)\rho(a), \tag{2.34}$$

since, if ρ vanishes on the generator of $\mathcal{R}'$, it vanishes on $\mathcal{R}'$ itself. Furthermore, in (2.34) we have adopted the notation $A=a+\mathcal{R}'$, $B=b+\mathcal{R}'$.

Equation (2.34) almost guarantees that our mapping has been constructed. To arrive at the special form (2.29), which is particularly useful for the applications we have in mind, we define the mapping σ on a $U(\lambda,\mu)$ monomial A by

$$\sigma(A)=\rho(A)\,(\lambda-\mu)^{(n-1)},\quad n=1,2,3\cdots, \tag{2.35}$$

where n is a *weight* associated with A in the following manner: As a monomial of $U(\lambda,\mu)$, A should be expressible in the form

$$A=a_1\times a_2\times\cdots\times a_n, \tag{2.36}$$

where the $a_j\in U(\lambda,\mu)$, $j=1,\ldots,n$, cannot be further reduced by $U(\lambda,\mu)$ products. Note that we have omitted to show in the last relation the specific association in the (nonassociative) product which makes up A; this is immaterial for our purposes. The point is that n factors enter the expansion of A and that number is the weight of A.

Now, in (2.36), the factors a_j, $j=1,\ldots,n$, which cannot be further reduced are either of the form

$$a_k=l_k+\mathcal{R}',\quad l_k\in\mathcal{L}, \tag{2.37a}$$

or of the form

$$a_i=\alpha_i+\mathcal{R}',\quad \alpha_i\in F. \tag{2.37b}$$

It can easily be shown, now, that the choice (2.35) is consistent with (2.29). Thus, let n and m be the weights of monomials A and B, respectively. Then,

$$\begin{aligned}\sigma(A\times B)&=\rho(A\times B)(\lambda-\mu)^{n+m-1}\\&=\lambda\rho(a)\rho(b)\,(\lambda-\mu)^{n+m-2}+\mu\rho(b)\rho(a)\,(\lambda-\mu)^{n+m-2}.\end{aligned} \tag{2.38}$$

On the other hand,

$$\begin{aligned}&\lambda\sigma(A)\,\sigma(B)+\mu\,\sigma(B)\,\sigma(A)\\&\quad=\lambda\rho(a)\,(\lambda-\mu)^{(n-1)}\rho(b)(\lambda-\mu)^{(m-1)}\\&\qquad+\mu\rho(b)\,(\lambda-\mu)^{(m-1)}\rho(a)\,(\lambda-\mu)^{(n-1)},\end{aligned} \tag{2.39}$$

where we have used the result $\rho(A)=\rho(a)$ since ρ vanishes on $\mathcal{R}'$.

Obviously, the right-hand sides of (2.38) and (2.39) are identical and the desired mapping has been constructed.

Several remarks are now in order. To begin with more trivial ones, let us make it clear that all applications we have in mind involve $U(\lambda,\mu)$ monomials of weight 2. Consequently, we shall gear all specific examples onto such cases. Secondly, we want to stress that our restriction to monomials in the proof of Theorem 2.2 involves no loss of generality since any $U(\lambda,\mu)$ element can be expressed as a sum of monomials. Consequently, the only complication involved is that one has to deal, separately, with more than one product between monomials.

The nature of $\mathcal{A}_{\lambda\mu}$ as an associative algebra of $\mathcal{L}$ monomials becomes evident from the mapping σ introduced during the proof of Theorem 2.2. Thus, consider, e.g., the $U(\lambda,\mu)$ element A of the form $a+\mathcal{R}'$, $a\in\mathcal{L}$. Then, since $n=1$,

$$\sigma(A)=\rho(A)=\rho(a) \tag{2.40}$$

and $\sigma(A)\,(\in\mathcal{A}_{\lambda\mu})$ can now be associated with the Lie algebra element a through ρ. This argument can be extended to monomials of higher weight. Thus, e.g., $\rho(l_1)\rho(l_2)$ in (2.32) is an $\mathcal{A}_{\lambda\mu}$ product of the Lie elements l_1 and l_2. Finally, for $\alpha\in F$ we have $\rho(\alpha)=\alpha$, i.e., α becomes an $\mathcal{A}_{\lambda\mu}$ scalar.

Note, now, a subtle difference between $\mathcal{A}_{\lambda\mu}$ and $\mathcal{A}$ as algebras of $\mathcal{L}$ monomials. In the case of $\mathcal{A}$, it is the projection mapping that associates the Lie element l, say, to the enveloping element $l+\mathcal{R}$. Thus, the as-

sociation is always made modulo the ideal $\mathcal{R}$. This fact accounts for the permissibility of rearranging monomials in $\mathcal{A}$ (PBW) theorem). On the other hand, once we are in $\mathcal{A}_{\lambda\mu}$ no rearrangements can be made (e.g., $l_1 l_2$ and $l_2 l_1$ are unrelated) as the ideal $\mathcal{R}'$ cannot be used to effect such rearrangements (the mapping ρ vanishes identically on $\mathcal{R}'$). Put in another way, there can be no PBW-like theorem for $\mathcal{A}_{\lambda\mu}$. To appreciate this statement consider the following question: Why can one not work with $\mathcal{T}(\mathcal{L})$ $[\mathcal{T}'(\mathcal{L})]$ exclusively instead of with $\mathcal{A}(\mathcal{U}(\lambda,\mu))$ and, thus, encounter Lie elements directly? The fact is that without $\mathcal{R}(\mathcal{R}')$ no reordering would have been meaningful. More specifically, a term such as $e_1 \otimes e_2$ would be completely different from $e_2 \otimes e_1$; these two expressions could not have been related in any manner. The PBW theorem constitutes the culmination of the role of $\mathcal{R}(\mathcal{R}')$ in the formation of the enveloping algebra (UEMA) of a Lie algebra. In this respect, $\mathcal{A}_{\lambda\mu}$ is like $\mathcal{T}(\mathcal{L})$, i.e., no rules pertaining to rearrangements can be established for it.

According to what we have said above, both $\mathcal{A}$ and $\mathcal{A}_{\lambda\mu}$ can be thought of as algebras of $\mathcal{L}$ monomials. Extending our considerations to $\mathcal{U}(\lambda,\mu)$, we see that the projection mapping makes it possible to identify, e.g., L_j and l_j where $L_j = l_j + \mathcal{R}'$, $l_j \in \mathcal{L}$. Similarly, we can write $l_1 \times l_2$ for $L_1 \times L_2$ and call $l_1 \times l_2$ a nonassociative Lie monomial belonging to $\mathcal{U}(\lambda,\mu)$. Theorem 2.2 finally permits us to write

$$l_1 \times l_2 = \lambda l_1 l_2 + \mu l_2 l_1. \tag{2.41}$$

We shall take (2.41) to reflect the practical essence of Theorem 2.2 for monomials of weight 2. Our considerations can be extended to monomials of larger weight with the resulting formulas being, of course, more complicated. Note an important consequence:

$$\alpha \times \beta = (\lambda + \mu)\,\alpha\beta \quad (\alpha, \beta \in F), \tag{2.42}$$

where by $\alpha \times \beta$ we mean $(\alpha + \mathcal{R}') \times (\beta + \mathcal{R}')$. From (2.42), the reason for our distinguishing the degree from the weight of a $\mathcal{U}(\lambda,\mu)$ monomial becomes obvious. Thus, scalar factors in a given $\mathcal{U}(\lambda,\mu)$ monomial do contribute a term $(\lambda - \mu)$ in the denominator of (2.34) which is compensated by the factor $(\lambda - \mu)$ entering through (2.35). The monomial $(\alpha_1 + \mathcal{R}') \times (l_1 + \mathcal{R}') \times (l_2 + \mathcal{R}') \times (\alpha_2 + \mathcal{R}') \times (l_3 + \mathcal{R}')$, for example, has degree 3 and weight 5.

We shall use the result of Theorem 2.2, in particular through its more straightforward form (2.41), in both the next section and the appendix.

III. GENERAL PHYSICAL CONSIDERATIONS AND AN INTERESTING SPECIFIC EXAMPLE

The contact between Lie groups (algebras), on the one hand, and field theory, on the other, comes through representation theory. Basically, one wants to relate observed (or conjectured) symmetries to the structure of a Lie group. One then works with field variables that belong to a space on which the assumed symmetry (Lie) group acts through a representation. As is well known, the elements of the Lie algebra of a Lie group[13] correspond (via their representation) to the generators of infinitesimal transformations. Accordingly, algebra representations are not interesting enough by themselves. However, a relation between an element of the Lie algebra and an element of the Lie group is established via the so-called exponential mapping. The latter can be specified in an abstract topological manner, but its more practical and readily applicable form is obtained by referring to the iniversal enveloping algebra of $\mathcal{L}$. Secondly, the existence of the universal enveloping algebra of a Lie algebra $\mathcal{L}$ is of vital practical importance when it comes to representations of $\mathcal{L}$.

To appreciate the last statement consider what is usually done when one deals with a representation ρ of a Lie algebra $\mathcal{L}$ on some vector space. One automatically sets

$$\rho([l_1 l_2]) = \rho(l_1)\rho(l_2) - \rho(l_2)\rho(l_1). \tag{3.1}$$

Similarly, when one considers two successive infinitesimal transformations one assumes that, e.g., $\rho(l_2)\rho(l_1)$ makes sense as representing an l_1-transformation. In both the above examples, it is the framework of the universal enveloping algebra that needs to be employed; otherwise, the object $l_2 l_1$ would not make any sense. Fortunately, the homomorphism which always exists between $\mathcal{L}$ and $\mathcal{A}_{\mathcal{L}}$ ensures that (3.1) can, indeed, be used.

To be more precise, let us recall certain important aspects from the theory of representations of a universal enveloping algebra. In particular, we know that the set of all representations of $\mathcal{L}$ on a vector space V is in one-to-one correspondence with the set of all representations of $\mathcal{A}$ on V ($\mathcal{A}$ denotes the universal enveloping algebra of $\mathcal{L}$)[14].

Note that, as an associative algebra, $\mathcal{A}$ has a faithful representation by linear transformations on a certain vector space. It follows that any Lie algebra also has a faithful representation by linear transformations. This is a result which is often used in connection with local field theory.

Next, let us briefly comment on the connection between the enveloping algebra and the Lie group. As is well known, the exponential mapping relates an element X of the Lie algebra $\mathcal{L}$ to the element $\exp X$ of G.[15] In its abstract definition the mapping $\exp: \mathcal{L} \to G$ sends lines in $\mathcal{L}$ (i.e., vectors along a given direction) to geodesic curves on the group manifold. However, a more practical approach, once again, is arrived at if we refer to the enveloping algebra and think of $\exp X$ in the conventional way, i.e.,

$$\exp X = I + X + \frac{1}{2!}X^2 + \frac{1}{3!}X^3 + \cdots. \tag{3.2}$$

Now, (3.2) really makes sense if $\exp X$ stands for a representation of the corresponding group element and I, X, X^2, etc., are representations of corresponding enveloping algebra elements. Then, we can interpret (3.2) as follows: $\exp X$ is a (finite) group transformation which is the sum of a series of infinitesimal transformations generated by X.

The general conclusion from the above discussion is that the universal enveloping algebra $\mathcal{A}$ of a Lie algebra

is of indispensable practical value when it comes to Lie representations. Furthermore, the framework of $\mathcal{A}$ is necessary for expanding the action of a group element in terms of a series of infinitesimal transformations. In the case of a UEMA of a Lie algebra, one does not know whether the parallel to a group, standing in relation to $\mathcal{U}(\lambda,\mu)$ as G stands in relation to $\mathcal{A}$, exists and, if it does, what its mathematical structure is. Clearly, if one wants to see a $\mathcal{U}(\lambda,\mu)$ mathematical scheme finding applications in physics one would have to have the parallel of the Lie group. Since, now, we have built $\mathcal{U}(\lambda,\mu)$ from the Lie algebra $\mathcal{L}$, one may be tempted to think that perhaps the group structure of G can support $\mathcal{U}(\lambda,\mu)$ as its nonassociative, (λ,μ)-parametrized enveloping algebra. That this is not so, can be seen very easily if we consider the Lie group product

$$g_1 \cdot g_2 = \exp tX \cdot \exp tY. \tag{3.3}$$

The well-known Campbell—Hausdorf formula gives

$$g_1 \cdot g_2 = \exp\ t(X+Y) + \frac{t^2}{2}[X,Y] + \mathcal{O}(t^3)\ , \tag{3.4}$$

where $\mathcal{O}(t^3)$ contains third-order terms in t. For convenience let us assume $|t| \ll 1$ so that $\mathcal{O}(t^3)$ will contain negligible terms. In (3.4) $[X,Y]$ stands for the enveloping algebra expression $XY - YX$. If, now, we substitute the UEMA framework for the universal enveloping algebra without changing the underlying group, we face difficulties. In particular, the Lie bracket $[X,Y]$ in (3.4) will be substituted by[16] $[X,Y]' = (\lambda-\mu)[X,Y]$. However, such a redefinition shifts the product $g_1 \cdot g_2$ (see Fig. 1). Indeed, $t(X+Y) + \frac{1}{2}t^2(\lambda-\mu)[X,Y]$ will not lie along the direction of $t(X+Y) + \frac{1}{2}t^2[X,Y]$ in the vector space $\mathcal{L}$. Accordingly, the exponential curve of the frist expression will not coincide with that of the second on the group manifold. Hence, to say the least, the adoption of the UEMA in the place of the universal enveloping algebra of $\mathcal{L}$ alters the group structure of the underlying manifold.

Arguing in a cruder way, we can say that the UEMA of a Lie algebra $\mathcal{L}$, even though constructed from the latter, does not go naturally with $\mathcal{L}$ or with the Lie group G. It seems reasonable that one should try to find (non-Lie) structures $\mathcal{L}'$ and G' which should be related to $\mathcal{L}$ and G, respectively, in some simple way—much as $\mathcal{U}(\lambda,\mu)$ is related to $\mathcal{A}$. Only then would one have a completely non-Lie framework which, in principle, can be employed in place of the customary Lie framework. In such a case, the parameters λ and μ would have to be given some physical interpretation. We shall make some relevant speculations on this matter in the next section.

We shall not, in this paper, attempt to discover the parallel of G for a UEMA. Instead, we shall turn our attention to representations of the UEMA itself. We shall think of them as infinitesimal (non-Lie) transformations. Since a UEMA is a nonassociative algebra, we do not expect that it has linear representations on a vector space. Let us, in fact, recall the requirements for a representation to be linear:

(1) $(A+B)x = Ax + Bx$,

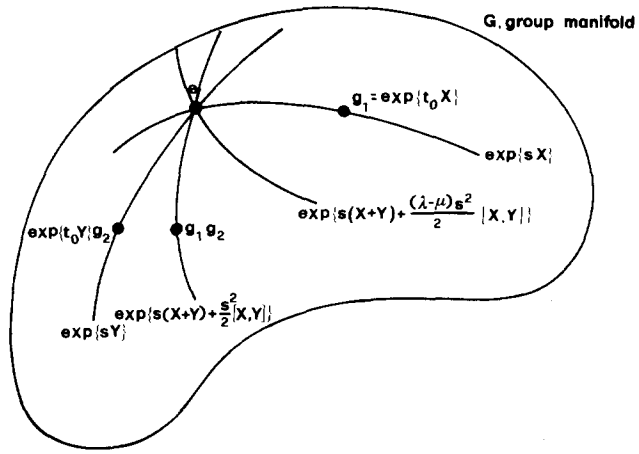


FIG. 1. The geodesic $\exp\{s(X+Y) + \frac{1}{2}s^2[X,Y]\}$, on the group manifold, is shifted if $\lambda - \mu \neq 1$.

(2) $(\alpha A)x = A(\alpha x)$,

(3) $(AB)x = A(Bx)$,

where x belongs to the representation space, A and B are representations of algebra elements, and α belongs to the field of scalars over which both the algebra and the representation vector space are defined. It is not hard to see that the last property is not shared by the UEMA. In fact, let $a, b \in \mathcal{L}$ with A, B being their respective representations. Then, we set $Bx = y$ and $A(Bx) = Ay$. On the other hand, $A \times B = \lambda AB + \mu BA$ from which follows $(AB)x = \lambda Ay + \lambda B(Ax)$. This expression is not, in general, equal to Ay. It follows that the representations of a UEMA are not linear.

We devote the rest of this section to an application of interest and, for this purpose, we restrict ourselves to the Lie group $U(1)$. We recall that local $U(1)$ transformations, within the conventional Lie framework, lead to the introduction of an electromagnetic field potential[17] in a very natural way. To be specific, suppose we have two charged scalar fields, ϕ and ϕ^*, in a theory which is invariant under local $U(1)$ transformations. The electromagnetic field potential A_μ is then introduced in order to covariantize the derivative ∂_μ. The Lagrangian density

$$\mathcal{L} = -\ \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} - \nabla_\mu \phi\, \nabla^\mu \phi^* - V(\phi\phi^*), \tag{3.5}$$

where

$$\nabla_\mu \phi = (\partial_\mu - ieA_\mu)\phi, \quad \nabla_\mu \phi^* = (\partial_\mu + ieA_\mu)\phi^*,$$
$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$$

and $V(\phi\phi^*)$ is a polynomial expression in $\phi\phi^*$, is invariant under gauge transformations of the second kind

$$\phi \to \exp[i\alpha(x)]\phi, \quad \phi^* \to \exp[-i\alpha(x)]\phi^*, \tag{3.6a}$$

$$A_\mu \to A_\mu - \frac{1}{e}\partial_\mu \alpha(x). \tag{3.6b}$$

In particular, (3.6a) implies that $\phi\phi^* \to \phi\phi^*$.

Suppose, now, that we choose to work within the non-

Lie framework of a UEMA. (We re-emphasise that, since we lack the full mathematical machinery, we confine ourselves to infinitesimal transformations.) A new possibility now arises. We start with the Lagrangian density

$$\mathcal{L} = -\tfrac{1}{2}\partial_\mu \phi\, \partial^\mu \phi^* - \tfrac{1}{2}m^2\phi\phi^*, \tag{3.7}$$

i.e., we are specializing to the simple case $V(\phi\phi^*) = \frac{1}{2}m^2\phi\phi^*$. If we want our theory to be invariant under $U(1)$ related local transformations and we work within the framework of the UEMA of $U(1)$, we could consider the following infinitesimal transformations:

$$\phi \to \exp[i\alpha(x)]\,\phi \approx \phi + i\alpha(x)\phi, \tag{3.8a}$$

$$\phi^* \to \exp[-i\beta(x)]\phi^* \approx \phi^* - i\beta(x)\phi^*. \tag{3.8b}$$

We shall comment later on the phase difference in the way ϕ and ϕ^* transform.

Now, $\alpha(x)$ and $\beta(x)$ are actually multiplied by the generator of infinitesimal $U(1)$ transformations.[18] As is the usual practice, however, the aforementioned generator has not been denoted explicitly, since it can be scaled via $\alpha(x)$ [and $\beta(x)$] to the identity operator [due to the one-dimensionality of $U(1)$]. But, whereas the identity operator of the usual (Lie-framework) representation has the trivial property

$$(\rho(I)\rho(I))\varphi = \rho(I)\phi = \phi, \tag{3.9}$$

within the UEMA framework we have

$$\rho(I)\times\rho(I) = \rho(I\times I) = \rho((\lambda+\mu)I) = (\lambda+\mu)\rho(I). \tag{3.10}$$

We see, then, that (neglecting higher order terms in the infinitesimals) one now obtains

$$\begin{aligned}\phi\phi^* &\to (\lambda+\mu)\phi\phi^* + i(\lambda+\mu)(\alpha-\beta)\phi\phi^* \\ &= (\lambda+\mu)[1+i(\alpha(x)-\beta(x))]\,\phi\phi^*\end{aligned} \tag{3.11}$$

Consider, first, the case $\alpha \equiv \beta$. One sees that the term $\frac{1}{2}m^2\phi\phi^*$ in (3.7) must be replaced by $[m^2/2(\mu+\lambda)]\phi\phi^*$. At the same time, one has to make the replacement $\partial_\mu \to \nabla_\mu$ where

$$\nabla_\mu\phi = \frac{1}{\sqrt{\mu+\lambda}}\,(\partial_\mu - ieA_\mu)\phi,$$

$$\nabla_\mu\phi^* = \frac{1}{\sqrt{\mu+\lambda}}\,(\partial_\mu + ieA_\mu)\phi^*. \tag{3.12}$$

Thus, one obtains the regular formalism of electrodynamics with a redefinition of the matter (scalar) fields.

Suppose, on the other hand $\alpha - \beta = \gamma \neq 0$. One could still obtain an invariant $\phi\phi^*$ term by choosing

$$(\lambda+\mu)(1+i\gamma(x)) = 1, \tag{3.13}$$

or, for infinitesimal γ,[19]

$$\lambda + \mu = \exp(-i\gamma). \tag{3.14}$$

It is straightforward to see that the same factor, i.e., $(\lambda+\mu)\exp(i\gamma)$, will appear in front of $\nabla_\mu\phi\nabla^\mu\phi^*$. By (3.14) this factor will, once again, disappear. The big difference is that we now need *two* gauge fields. In particular, we get

$$\nabla_\mu\phi = (\partial_\mu - iaA_\mu)\phi, \tag{3.15}$$

where A_μ transforms according to

$$A_\mu \to A_\mu - \frac{1}{a}\,\partial_\mu\alpha(x), \tag{3.16}$$

and

$$\nabla_\mu\phi^* = (\partial_\mu + ibB_\mu)\phi^*, \tag{3.17}$$

where B_μ transforms according to

$$B_\mu \to B_\mu - \frac{1}{b}\,\partial_\mu\beta(x). \tag{3.18}$$

Even though the above scheme is quite unorthodox, it bears striking similarity to the (phenomenological) one recently proposed by Hsu[8] in his version of superweak interaction interpretation of CP violation by the K^0 system. For example, the vector–scalar part of Hsu's Lagrangian reads

$$\begin{aligned}\mathcal{L}_{\text{Hsu}} = &-(\partial_\mu + iF^*(A_\mu - i\beta C_\mu))K^{0*}(\partial_\mu - iF(A_\mu + i\beta C_\mu))K^0 \\ &- m^2K^{0*}K^0 - \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} - \tfrac{1}{4}C_{\mu\nu}C^{\mu\nu} \\ &+ (\text{terms pertaining to choice of gauge}),\end{aligned} \tag{3.19}$$

where F is a complex and β a real number and the meanings of $F_{\mu\nu}$ and $C_{\mu\nu}$ are evident. Furthermore, A_μ and C_μ transform according to

$$A_\mu \to A_\mu + F_{\text{real}}\,\frac{\partial_\mu\Lambda(x)}{|F|^2}, \quad C_\mu \to C_\mu - F_{im}\,\frac{\partial_\mu\Lambda(x)}{\beta|F|^2}. \tag{3.20}$$

By rearranging our A_μ and B_μ one should obtain Hsu's Lagrangian with the redefined vector fields obeying (3.20). Note that (3.20) implies that our γ is constant, i.e., that $\alpha(x)$ and $\beta(x)$ are functions differing by a constant.

Hsu's model actually requires the presence of spin-$\frac{1}{2}$ fields also, but no added difficulty really enters by introducing such fields. Once more, we require that $\psi(x)$ and $\bar{\psi}(x)$ have the transformation properties

$$\psi \to \exp[i\alpha(x)]\psi, \quad \bar{\psi} \to \exp[-i\beta(x)]\,\bar{\psi} \tag{3.21}$$

and the spinor sector of Hsu's Lagrangian follows in the same manner.

The drawback from all this is that we have restricted ourselves to infinitesimal transformations since we lack the existence of the analog to a group structure. Still, it should become evident from this example that there could, indeed, be merits in the whole UEMA scheme.

We close our present discussion with some remarks on the phase difference $\gamma(=\alpha-\beta)$ between the way the particle transformation and that of its charge conjugate [e.g., relations (3.8a, b)]. Presumably such a phase difference should be connected with a "charge" violation which Hsu already suggests in Ref. 8. To follow Hsu's reasoning, his theory suggests that the charge of all particles fulfills the relation

$$\text{"charge"} = q + \lambda S, \tag{3.22}$$

with λ a very small (constant) parameter, S the strangeness and q standing for $\pm e$ or 0. Furthermore, the "charge" of (3.22) is conserved implying also conservation of strangeness.

Most probably the relation between Hsu's additional charge and our charges a, b should be something like

$$(a-b)/2=\lambda S. \tag{3.23}$$

IV. DISCUSSION AND SPECULATIONS

As we have already mentioned in the introduction, many speculations can be made concerning the use of two parameters one has in hand. We note that, for our case, the two parameters are presumably connected with a general non-Lie symmetry scheme which, hopefully, can be employed for physical descriptions. In the previous section, we have indicated that the corresponding (non-Lie) framework is far from being complete. On the other hand, we have also shown, in Sec. III, that one unorthodox scheme which can be deduced from UEMA considerations bears striking resemblance to a phenomenological theory proposed for the explanation of a very important problem in elementary particle physics. Accordingly, we are encouraged to proceed and comment briefly on a few possibilities of UEMA relevance to physics even though we are aware of the fact that a complete UEMA formalism is not as yet fully developed.

Our most immediate thoughts turn to current algebra. Indeed, some recent approaches aimed towards the better understanding of symmetry breaking may have a natural position within the realms of the UEMA of $SU(3)$. We are referring to the deformed algebra approach of Oehme[20] and Mathur,[21] as well as the closely related "weighted" $SU(3)$ transformations of Rest and Welling.[22] These attempts lead to insights concerning the Cabibbo angle as well as several good predictions connected with K_{l_3} form factors. We shall indicate briefly that there exist possibilities of incorporating such approaches into the UEMA scheme. In passing, we would like to remark that what we hope the relevance of the UEMA machinery will be, once fully developed, is precisely its capacity to incorporate apparent deviations from an ideally perfect symmetry within a background where the aforementioned symmetry firmly resides. In this sense, the deformed algebra ideas, for example, constitute exactly the kind of approach that fits into the ideas behind what we hope to be the physical relevance of the UEMA scheme. Perhaps the most commonplace term "hidden symmetry", which is also becoming of use, best reflects the content of the last two statements. [Indeed, it may be appropriate to attribute the ideal symmetry, which always resides in the background, to some hidden (particle?) structure of unobservable fields and reserve the somewhat looser symmetry framework (e.g., UEMA) to reflect what is observed.]

So as not to make our discussion too involved we shall concentrate, for the most part, on Oehme's scheme[20] which is the simplest of the three we have mentioned above. Consider a chiral octet of $SU(3)$ currents J_μ^a $(=V_\mu^a+A_\mu^a)$ whose charges obey chiral $SU(3)\times SU(3)$ commutation rules at equal times (ET). Suppose we believe the chiral $SU(2)\times SU(2)$ part of the symmetry to be exact ($m_\pi=0$ and absence of electromagnetic effects). According to Oehme, a very simple way to implement this state of affairs is to redefine the currents by

$$J_\mu^a \to J_\mu^{a\prime}=\begin{cases}\xi J_\mu^a, & a=4,5,6,7\\ J_\mu^a, & a=1,2,3,8.\end{cases} \tag{4.1}$$

Accordingly, equal-time charge commutators remain the same except for those among charges with indices running between 4 and 7. In particular,

$$_{ET}[Q_k', Q_j']=if_{kjl}\xi^2 Q_l', \quad 4\le k,j\le 7, \tag{4.2}$$

i.e., a factor of ξ^2 enters on the right-hand side. This procedure makes it possible to go ahead with a computation of the Cabibbo angle and, more important, to obtain some insight as to its meaning.

Now, (4.2) induces some thoughts in connection with the UEMA formalism of $SU(3)$. Suppose, in general, that we deviate from a Lie to the corresponding UEMA framework in a minimal manner, by which we mean that we use $\mathcal{U}_L(\lambda,\mu)$ to replace $\mathcal{A}_L$. But whereas $\mathcal{A}_L$ can be put into natural correspondence (homomorphism) with the Lie algebra $\mathcal{L}$, $\mathcal{U}_L(\lambda,\mu)$ corresponds—in the same manner—to the Lie algebra $(\lambda-\mu)\mathcal{L}$. By $(\lambda-\mu)\mathcal{L}$ we mean the Lie algebra $\mathcal{L}'$ with elements formally identical to those of $\mathcal{L}$ but with Lie product

$$[l_1 l_2]'=(\lambda-\mu)[l_1 l_2]. \tag{4.3}$$

It is not immediately obvious from (4.3) whether the parameter $\alpha\equiv\lambda-\mu$ is of any special significance. On one hand, (4.3) looks like a redefinition of the basis elements of $\mathcal{L}$ or, alternatively, of its structure constants. Thus, the $SU(3)$ equal-time commutator

$$_{ET}[Q_a, Q_b]=if_{abc}Q_c \tag{4.4}$$

goes into

$$_{ET}[Q_a, Q_b]'=i\alpha f_{abc}Q_c \tag{4.5}$$

with a,b,c running through all $SU(3)$ indices. In turn, (4.5) is equivalent to

$$_{ET}[Q_a', Q_b']=i\alpha f_{abc}Q_c', \tag{4.6}$$

where $Q_j'=\alpha Q_j$, $j=1,\ldots,8$.

On the other hand, the study of subalgebras—or, more generally, of inner structures—of UEMA may open up possibilities of the kind described by (4.2). For example, the subalgebra $SU(2)$ of $SU(3)$ has its corresponding $(\lambda-\mu)SU(2)$ image. Suppose, then, we build a (λ,μ) generalization of $SU(3)$ in two stages, i.e., we first construct $(\lambda'-\mu')SU(2)$ and then, through a second generalization $(\tilde\lambda,\tilde\mu)$, we introduce the full (λ,μ) generalized $SU(3)$. In that case, we would have two parameters, $\alpha'=\lambda'-\mu'$ and $\tilde\alpha=\tilde\lambda-\tilde\mu$, which could be used to obtain commutation relations such as those given by (4.2). In fact, if it is our contention (and/or ambition) that the (λ,μ) parametrization measures deviations from exact symmetries, we would (physically) expect a different (λ,μ) content for the $SU(2)\times SU(2)$ part of the algebra of currents than for the whole of $SU(3)\times SU(3)$—as observation warrants. Suppose, then, that (4.5) is used to describe the full $SU(3)$ part of the algebra of currents, whereas for the isospin restriction we accept a generalization (λ',μ') closer to $(1,0)$ and reflected by the relations

$$_{ET}[R_a, R_b]' = i\epsilon_{abc}\alpha' R_c, \quad 1 \leq a, b \leq 3. \tag{4.7}$$

In that case, Q_1, Q_2, and Q_3 would not be the true isospin charges (R_j). Rather, there should be a relation between the two sets given by

$$Q_k = \beta R_k, \quad k = 1, 2, 3, \tag{4.8}$$

where $\beta = \alpha/\alpha'$.

Similarly, if the (λ, μ) generalization of the $U(1)$ symmetry associated with the hypercharge was effected through (λ'', μ'')—presumably closer to (1, 0) than (λ, μ)—then Q_8 would be related to the hypercharge charge Y by

$$Q_8 = \gamma Y, \tag{4.9}$$

where $\gamma = \alpha/\alpha''$.

In particular, Oehme's choice is $\alpha' = \alpha'' = 1$ and $\beta = \gamma = \alpha$. Mathur's work,[21] on the other hand, deals with $1 \neq \alpha' \neq \alpha'' \neq 1$. With some minor adjustments one can also accommodate, along the same lines, the "weighted" $SU(3)$ transformation approach of Rest and Welling.[22]

As a final speculation, let us take note of the fact that intrinsic symmetries seem to form a hierarchy [e.g., $U(1)$, $SU(2)$, $SU(3)$] whereby the more complex the intrinsic symmetry the more it appears to be broken. As we have already argued in this section, the UEMA parametrization (λ, μ) can conceivably measure deviations from a perfect (Lie) symmetry through its departure from (1, 0). Consider, now, for a moment, the ideas of Lee and Wick,[23] on one hand, and the work of Kirzhnits and Linde[24] and of Weinberg,[25] on the other. What emerges from these studies is the mutability of physical laws, or, perhaps more appropriately, the existence of domains in the universe which may exhibit different characteristics of symmetry behavior. Suppose that λ and μ are taken to be local functions of space–time (presumably very slowly varying). Then, one can think of a space–time "evolution" of intrinsic symmetries—an idea which comes close to the thinking of Kirzhnits and Linde as well as Weinberg. Perhaps, also, one could find a relation between, say, $(\mu - 0)$ and the nonvanishing expectation value $\langle\phi(x)\rangle$ in an abnormal domain of Lee and Wick.

It would carry us too far to continue such speculations. In fact, it would be presumptuous to go on doing so when a complete non-Lie framework for symmetry descriptions is not as yet available. It would be interesting, in any case, to pursue unorthodox schemes such as the one we have introduced here. For the particular case of a UEMA the important next step is to look for the corresponding "mutation" group.

ACKNOWLEDGMENTS

The author is especially indebted to Professor G. Furlan whose insights have helped widen his point of view considerably. In addition, several communications from Professor R. M. Santilli have been extremely helpful in putting some matters in perspective. Finally, it is a pleasure to thank Professor Abdus Salam, the International Atomic Energy Agency and UNESCO for hospitality at the International Centre for Theoretical Physics, Trieste, Italy.

APPENDIX

In this appendix we shall be concerned with an application of the PBW theorem for a UEMA, proved in Sec. II, pertaining to monomials of degree 2. For this particular case the $\mathcal{U}(\lambda, \mu)$ expansion does not contain any $\mathcal{A}$ terms.

The most practical implication of the PBW theorem, in connection with universal enveloping algebras is the following. Let $\{e_1, \ldots, e_n\}$ be an ordered basis for $\mathcal{L}$. Then, the element 1 together with all standard monomials $e_{j_1} e_{j_2} \cdots e_{j_k}$, $j_1 \leq j_2 \leq \cdots \leq j_k$, form a basis for the universal enveloping algebra of $\mathcal{L}$. This remark leads, naturally, to a Gell-Mann–Okubo mass formula for, say, an $SU(3)$ octet [i.e., $\mathcal{L} = SU(3)$]. Indeed, the well-known formula

$$T^\mu_\nu = a\delta^\mu_\nu + bA^\mu_\nu + c\sum_j A^\mu_j A^j_\nu, \tag{A1}$$

where T^μ_ν is an $SU(3)$ tensor operator (of second rank), is a natural consequence of the fact that a general monomial of degree 2 can be expanded as a linear combination of standard monomials of degree 2 and less. Thinking, in particular, in terms of a matrix representation of $\mathcal{L}$, it follows that

$$AB = aI + \sum b_j E_j + \sum c_{j_k j_l} E_{j_k} E_{j_l}, \quad j_k \leq j_l, \tag{A2}$$

where A, B, as well as the E_j, are matrix representations of Lie elements (the E_j, in particular, correspond to matrix representations of the basis of $\mathcal{L}$). Not thinking in terms of the universal enveloping algebra $\mathcal{A}$, AB corresponds to a second-rank [$SU(3)$] tensor. Accordingly, (A1) is a tensorial version of (A2).

We want to derive the analog of the Gell-Mann–Okubo mass formula within the framework of the UEMA of $SU(3)$. To this effect we shall neglect mixing effects between multiplets (even though such a mixing is very pronounced in some cases, e.g., $\phi - \omega$ mixing). Let us, following Okubo, confine our considerations to an $SU(3)$ octet (baryon octet for simplicity) and assume that the mass operator transforms as an $SU(3)$ scalar plus a $Y = T = 0$ of an $SU(3)$ octet. In making this assumption we have excluded the possibility of mixing between particles belonging to different representations of $SU(3)$ as already remarked. Choosing appropriate $SU(3)$ generators[26] A^i_j, $i, j = 1, 2, 3$, with $\sum_{j=1}^3 A^j_j = 0$—corresponding to the E_j of Eq. (A2)—it follows that T^3_3 ($T = Y = 0$) has the form

$$T^3_3 = aA^3_3 + b(A^3_j A^j_3 - \tfrac{1}{3}C_2), \tag{A3}$$

where C_2 is the second-order Casimir operator for $SU(3)$. One can obtain the Gell-Mann–Okubo mass formula from (A3) in a straightforward manner.

Returning now to our UEMA, the expansion (A2) is replaced by an expansion in terms of standard monomials with respect to $\otimes'$. Let us choose for $SU(3)$ the same basis as that leading to (A3). We must first make some adjustments. The standard monomial $A^3_j A^j_3$ will

be replaced by

$$A_j^3 A_3^j \to A_j^3 \times A_3^j = \lambda A_j^3 A_3^j + \mu A_3^j A_j^3. \tag{A4}$$

Using the commutation relation for the A_j^i,

$$[A_j^i, A_l^k] = \delta_l^i A_j^k - \delta_j^k A_l^i, \tag{A5}$$

we obtain

$$A_j^3 \times A_3^j = (\lambda + \mu) A_j^3 A_3^j + \mu (A_1^1 + A_2^2) - 2\mu A_3^3. \tag{A6}$$

Substituting $A_1^1 + A_2^2 = -A_3^3$, we finally have

$$A_j^3 \times A_3^j = (\lambda + \mu) A_j^3 A_3^j - 3\mu A_3^3. \tag{A7}$$

Next we consider the Casimir operator C_2 which is given by

$$C_2 = \tfrac{1}{2} A_j^i A_i^j. \tag{A8}$$

The replacement of (A8) for the case of the UEMA is

$$C_2 \to C_2' = \tfrac{1}{2}(\lambda A_j^i A_i^j + \mu A_i^j A_j^i) = (\lambda + \mu) C_2. \tag{A9}$$

Obviously, since λ and μ belong to the field of scalars F over which the algebra of $SU(3)$ is defined as a vector space, $(\lambda + \mu)C_2$ is an $SU(3)$ scalar inasmuch as C_2 is.

Recalling that[27]

$$A_3^3 = -Y, \quad A_j^3 A_3^j = C_2 - \mathbf{T}^2 + \tfrac{1}{2}Y^2 + 3/2Y, \tag{A10}$$

where $\mathbf{T}$ is the isospin and Y the hypercharge operator, we finally conclude that

$$T_3^3 = -aY + b[(\lambda + \mu)\{C_2 - \mathbf{T}^2 + \tfrac{1}{4}Y^2 + 3/2Y\} + 3\mu Y - \tfrac{1}{3}(\lambda + \mu)C_2], \tag{A11}$$

which gives

$$\Delta M = b\,\frac{2(\lambda + \mu)}{3}\,C_2 + [3b/2(\lambda + 3\mu) - a]Y + b(\lambda + \mu) \times [\tfrac{1}{4}Y^2 - T(T+1)]. \tag{A12}$$

Note that we recover the Gell-Mann–Okubo mass formula when $\lambda = 1$, $\mu = 0$.

A similar formula, not identical to the above, however, has been given[6] within the framework of the mutation algebra of an enveloping algebra, i.e., $\mathcal{A}(\lambda, \mu)$.

[1] P. Jordan, J. von Neumann, and E. Wigner, Ann. Math 35, 29 (1934).

[2] The usual way of regarding $\{a, b\}$, i.e., $\{a.b\} = ab + ba$, presupposes the existence of the product ab. Here, we want to remain quite general in that the product $\{a, b\}$ is considered to have been introduced *a priori* by itself.

[3] R. M. Santilli, Nuovo Cimento 51, 570 (1967); Suppl. Nuovo Cimento 6, 1225 (1968).

[4] A. A. Albert, Trans. Amer. Math. Soc. 64, 552 (1948).

[5] R. M. Santilli and G. Soliani, "Statistics and parastatistics formal unification," University of Torino, internal note (1967).

[6] R. M. Santilli, in *Analytical Methods in Mathematical Physics*, edited by R. P. Gilbert and R. G. Newton (Gordon and Breach, New York, 1970), Paper II-18; R. M. Santilli, Meccanica 1, 3 (1969); P. Roman and R. M. Santilli, Lett. Nuovo Cimento 2, 449 (1969).

[7] See N. Jacobson, *Lie Algebras* (Interscience, New York, 1962), p. 155.

[8] J. P. Hsu, Phys. Rev. D 9, 810 (1974); University of Texas, Austin, Preprint CPT 220 ORO-3992-176 (1974).

[9] S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).

[10] See p. 152 of Ref. 7.

[11] The association in the expression $a_1 \otimes' \cdots \otimes' a_n$ is arbitrary whenever not indicated [see comment after (2.8)].

[12] See p. 158 of Ref. 7.

[13] Generally speaking, a Lie algebra does not need the existence of a Lie group in order to be defined.

[14] The proof hinges on the following observation. If ρ is a representation of $\mathcal{L}$, we define a representation $\tilde{\rho}$ on $\mathcal{T}(\mathcal{L})$ which agrees with ρ on $\mathcal{L}$, i.e., $\tilde{\rho}(l) = \rho(l)$ for $l \in \mathcal{L}$. This representation is unique if we set, e.g., $\tilde{\rho}(l_1 \otimes l_2) = \rho(l_1)\,\rho(l_2)$, etc. Thinking, now, of a Lie algebra as the tangent space at the identity element e of the group manifold, the Lie product $[l_1 l_2]$ can be identified with $(L_1 L_2 - L_2 L_1)_e$ where L_1 and L_2 are vector fields over the manifold and L_{j_e} denotes the tangent vector at e corresponding to the vector field L_j. From these remarks it follows that

$$\tilde{\rho}(l_1 \otimes l_2 - l_2 \otimes l_1 - [l_1 l_2]) = \rho(l_1)\rho(l_2) - \rho(l_2)\rho(l_1) - \rho([l_1 l_2]) = 0.$$

In other words, $\tilde{\rho}$ vanishes on the ideal $\mathcal{R}$. Finally, the representation ρ is extended to a representation $\hat{\rho}$ of $\mathcal{A}$ by $\hat{\rho} \cdot \pi = \tilde{\rho}$, where π is the projection mapping of $\mathcal{T}(\mathcal{L})$ onto $\mathcal{A}[\equiv \mathcal{T}(\mathcal{L})/\mathcal{R}]$. Similar arguments hold for the converse route.

[15] More precisely, $\exp X$ belongs to that connected part of G which contains the identity element.

[16] This can be shown by explicit calculation involving exponential expansions within UEMA.

[17] R. Utiyama, Phys. Rev. **101**, 1597 (1956).

[18] Clearly, we can still talk about $U(1)$ generators since the UEMA contains (mod $\mathcal{R}'$) the elements of the Lie algebra of $U(1)$.

[19] Note that this relation implies a complex algebraic product. Explicitly, let us set $\lambda = 1$, $\mu = i\gamma$. Then we have, e.g., $A \times B = AB + i\gamma BA$ where A and B are (mod $\mathcal{R}'$) Lie algebra elements. We shall comment later on the possible meaning of γ.

[20] R. Oehme, Phys. Rev. D 7, 1162 (1973).

[21] V. Mathur, Phys. Rev. Lett. 29, 1419 (1972).

[22] J. Rest and D. J. Welling, Phys. Rev. D 3, 2783 (1971); D 5, 1983 (1972).

[23] T. D. Lee and G. C. Wick, Phys. Rev. D 9, 2291 (1974).

[24] D. A. Kirzhnits and A. D. Linde, Phys. Lett. B 42, 471 (1972).

[25] S. Weinberg, Phys. Rev. D 9, 3357 (1974).

[26] See p. 32 of P. Carruthers, in *Introduction to Unitary Symmetry* (Interscience, New York, 1966).

[27] See p. 151 of Ref. 26.

On the entire solutions of a functional equation in the theory of fluids

Fred Gross

Computer Science Laboratory, Naval Research Laboratory, Washington, D.C. 20375 and University of Maryland, Baltimore, Maryland

Charles F. Osgood and Chung-Chun Yang

Mathematics Sciences Staff, Naval Research Laboratory, Washington, D.C. 20375
(Received 27 May 1975)

In this paper, we shall give necessary and sufficient conditions for the existence of entire solutions to the functional equation $\psi^2+g\varphi^2=h$, where g and h are given nonzero polynomials in z. This functional equation arises when one studies the Percus–Yevick integral equation of hard sphere mixture. The contruction of all such entire solutions ψ and φ is presented. Also, we shall show that it is possible, in some cases, to prove the existence of solutions with prescribed asymptotic properties at $z=\infty$.

INTRODUCTION

Consideration of the Percus—Yevick integral equation of hard sphere mixture leads one to consider functional equations of the type

$$\psi^2+g\varphi^2=h \tag{1}$$

where g and h are given nonzero polynomials in z, and ψ and φ are required to be entire functions of z. In this paper we shall give necessary and sufficient conditions for the existence of entire solutions to (1). Special solutions of this type, for certain polynomials g and h, were obtained using singular integration techniques, by Penrose and Lebowitz.[1] We shall give a construction of *all* such entire solutions ψ and φ by using some different and relatively elementary arguments. Also we shall show that it is possible, in some cases, to prove the existence of solutions with prescribed asymptotic properties at $z=\infty$. (In one very special case it is even possible to show the existence of polynomial solutions and to give bounds upon the degrees of the solution polynomials ψ and φ.) Extensions where we replace the right-hand side of (1) by a more general function than a polynomial and (or) replace the left-hand side by a "norm form," of degree $n>2$ in n entire functions, look possible and are being investigated.

I. THEOREMS I–V AND LEMMA

Theorem I below is an existence theorem. Theorem II shows the existence of a convenient canonical form for those equations of type (1) actually having entire solutions. Theorems III and IV describe the general solution. Theorem V treats the asymptotic theory of solutions of equations of type (1).

Theorem I: There always exist entire solutions ψ and φ of (1), unless for some complex number a and positive integer α, $z-a$ divides h to the exact power $2\alpha-1$ while $(z-a)^{2\alpha}$ divides g.

The following definition is needed for the statement of Theorem II.

Definition: A functional equation

$$\psi^2+g_1\varphi^2=h_1, \tag{2}$$

where g_1 and h_1 are nonzero polynomials with g_1 dividing g and h_1 dividing h, will be said to be *constructively equivalent* to (1) if and only if one can construct an algorithm for producing two polynomials $\rho(z)$ and $\xi(z)$ such that whenever ψ_1 and φ_1 are entire solutions of (2), then $\psi=\rho(z)\psi_1$ and $\varphi=\xi(z)\varphi_1$ are entire solutions of (1) and, conversely, every pair of entire solutions of (1) is of the form $\rho(z)\psi_1$ and $\xi(z)\varphi_1$ for entire solutions ψ_1 and φ_1 of (2).

For the canonical representation we have:

Theorem II: Each equation of type (1) which actually has at least one pair of entire solutions is constructively equivalent to some equation of the form

$$\psi^2+qp\varphi^2=qq_1 \tag{3}$$

where p, q, and q_1 are relatively prime nonzero polynomials and q has no multiple zeros.

We next proceed to describe the structure of the general solution of (3).

Theorem III: Let us choose branches of $\sqrt{qp}$ and $\sqrt{qq_1}$. A pair of entire functions ψ and φ are a solution of (3) if and only if they are of the form

$$\psi=\sqrt{qq_1}\sin(\sqrt{qp}\,\gamma(z)),$$
$$\varphi=\sqrt{p^{-1}q_1}\cos(\sqrt{qp}\,\gamma(z))$$

where $\gamma(z)$ may be any multiple valued function which is analytic and single valued on the complex plane with suitable cuts from the zeros of pqq_1 to $z=\infty$ removed such that: (i) at each zero z_j of q, $\gamma(z)-n_j\pi(\sqrt{qp})^{-1}$ is analytic for some integer n_j, (ii) at each zero z_k of p, $\gamma-(l_k+\frac{1}{2})\pi(\sqrt{qp})^{-1}$ is analytic at $z=z_k$ for some integer l_k and (iii) at each zero z_r of q_1 of multiplicity m_r, $\gamma(z)+\frac{1}{2}t_r i\log(z-z_r)(\sqrt{qp})^{-1}$ is analytic, for some integer t_r, $-m_r\leq t_r\leq m_r$, such that $\frac{1}{2}(m_r+t_r)$ is an integer, (iv) given a solution pair ψ_1 and φ_1, and any associated γ, say γ_1, then γ_2 is another associated γ if and only if $\sqrt{qp}(\gamma_1-\gamma_2)=2n_1\pi$ for some integer n_1.

[Note if γ corresponds to the solution ψ_1, φ_1, then $\gamma+\pi(\sqrt{qp})^{-1}$ corresponds to the pair $-\psi_1$, $-\varphi_1$ and $-\gamma$ corresponds to the pair $-\psi_1$, φ_1.]

Theorem IV: Let $s(z)$ equal the product of the distinct zeros of $q_1(l)$. For each p, q, q_1, and sets of integers n_j, l_k, and t_r as in Theorem III there exists a function $\gamma_1(z)$ satisfying the conditions of Theorem III concerning $\gamma(z)$. Further, $\gamma_1(z)$ may be written in the form

$$\gamma_1(z) = (\sqrt{p(z)q(z)})^{-1} \int_a^z \sqrt{p(t)q(t)}(s(t))^{-1} f(t)\, dt,$$

for some polynomial $f(t)$ and some complex number a which is not a zero of pqq_1.

It is clear that any $\gamma(z)$ as in Theorem III, corresponding to a particular set of n_j, l_k, and t_r, and the $\gamma_1(z)$ in Theorem IV, corresponding to the same set of n_j, l_k, and t_r, differ by an entire function.

Theorem V: Set $\delta = \frac{1}{2}\deg(pq)$. Let us choose, arbitrarily, a polynomial F in z. (i) If $\delta > 1$ is an integer there exists some $\sqrt{pq}\,\gamma$ as in Theorem III such that, at $z=\infty$, $\sqrt{pq}\,\gamma$ is asymptotic to $F + O(z^{\delta-1})$. (ii) If $\delta > \frac{1}{2}$ is not an integer, there exists some $\sqrt{pq}\,\gamma$ as in Theorem III such that, at $z=\infty$, $\sqrt{pq}\,\gamma$ is asymptotic to $\sqrt{z}\,F + O(z^{\delta-1})$. (iii) If δ equals 0 or 1, then there exists some $\sqrt{pq}\,\gamma$ as in Theorem III such that, at $z=\infty$, $\sqrt{pq}\,\gamma$ is asymptotic to $F + O(\log z)$. If $\delta = \frac{1}{2}$ then there exists some $\sqrt{pq}\,\gamma$, as in Theorem III, such that, at $z=\infty$, $\sqrt{pq}\,\gamma$ is asymptotic to $\sqrt{z}\,F + O(1)$. (iv) If $\delta > 0$ and both $\sqrt{pq}\,\gamma_1$ and $\sqrt{pq}\,\gamma_2$ are asymptotic to either F or $\sqrt{z}\,F$ for some polynomial F, up to the error term given above for the appropriate value of δ, and if, further, γ_1 and γ_2 both give rise to the same constants n_j, l_k, t_r, then $\gamma_1 \equiv \gamma_2$. (v) In parts (i)–(iii) above, given any allowable set of values of the constants n_j, l_k, and t_r (see Theorem III), we may require also that the γ to be constructed there corresponds to these values of n_j, l_k, and t_r.

Lemma: If, for any two solutions ψ and φ of (1), $\sqrt{pq}\,\gamma$ is asymptotic to $\alpha\sqrt{z} + O(1)$ where $\alpha \neq 0$ is real then on any angular sector about $z=0$, not including the negative real axis, we see that $|\psi|$ and $|\varphi|$ are asymptotic to

$$\exp[\alpha\,|R\sqrt{z}\,| + O(|\log z|)], \quad \text{at } z=\infty.$$

Proof: Trivial.

Example: We now consider what we call a more general Penrose–Lebowitz equation, i.e.,

$$\psi^2 + (az^2+b)\varphi^2 = r(z^2),$$

where a and b are complex numbers, $a \neq 0$, $r(z^2) \not\equiv 0$ is a polynomial in z^2 with complex coefficients, and $(az+b)^2$ does not divide $r(z)$. We are required to find even entire solutions ψ and φ of the above equation, such that, for some real $\alpha \neq 0$, $|\psi|$ and $|\varphi|$ are each asymptotic to

$$\exp[\alpha\,|Rz\,| + O(|\log z|)],$$

on any angular sector about $z=0$ not containing either the positive or negative imaginary axis, at $z=\infty$.

Set $w=z^2$. By Theorem I we always have a solution. By Theorem V with $\delta = \frac{1}{2}$, and the Lemma, we have that for each integral choice of n_1 [or l_1 if $aw+b$ divides $r(w)$] and the allowable choices of t_r there are entire solutions $\psi_1(w)$ and $\varphi_1(w)$ with $|\psi_1(w)|$ and $|\varphi_1(w)|$ each asymptotic to

$$\exp[\alpha\,|R\sqrt{w}\,| + O(\log w)]$$

on any angular sector about $z=0$ which does not contain the negative real axis. Then $\psi_1(z^2)$ and $\varphi_1(z^2)$ satisfy the desired conditions.

From Theorem III(iv) and the comment after the statement of Theorem III, we see that if in the general Penrose–Lebowitz equation above the only free integral valued parameter in our solution is n_1(or l_1), then the only solutions are $\pm\psi$, $\pm\varphi$. This will be the case if q_1 is a nonzero constant. If q_1 is linear in z^2, then we have two choices of t_1, i.e., $t_1 = \pm 1$. Interchanging γ and $-\gamma$, if necessary, we may take $t_1 = 1$. Then by the comment after the statement of Theorem III, the uniqueness of ψ_1 and φ_1, up to sign, follows. If $q_1(w)$ has degree larger than one, we cannot have uniqueness, even up to sign, since for one choice of the t_r's we can make our $\psi(w) + \sqrt{aw+b}\,\varphi(w)$ and $q_1(w)$, and our $\psi(w) = \sqrt{aw+b}\,\varphi(w)$ and $q_1(w)$, each, have at least one common zero, while for another choice of the t_r's there are no common zeros in one of these two cases. [This becomes clear in the proof of Theorem III(iii).]

If we weaken the asymptotic condition slightly, there is not uniqueness. There exists an entire function $E_1(z)$[2] which is bounded on every ray out from zero but which is not a constant and is not even of finite order of growth. (The bound is not uniform, of course.) Adding such a function (we can construct many) to $(\sqrt{qp})^{-1}\gamma$, we would obtain $|\psi|$ and $|\varphi|$ asymptotic to

$$\exp[\alpha\,|R\sqrt{w}\,| + O(|\log w|)]$$

on each ray out from zero except the negative real axis, for a large collection of entire solutions ψ and φ.

If we have an equation analogous to the Penrose–Lebowitz equation in that both polynomials are even and the equation is already in the same form as (3), then it follows from Theorems III and V that if the coefficient of φ has degree in z^2 greater than two, it is impossible to ever satisfy the asymptotic condition (in the strong form) for all real $\alpha \neq 0$. At most, a countable number of different α's are possible. One would have to go through the proof of Theorem IV, in the particular case in order to hope to say more.

We note in passing that by Theorems III and V each equation of the form

$$\psi^2 + (az+b)\varphi^2 = r(z),$$

where $a \neq 0$ and b are complex numbers and $r(z)$ is a nonzero polynomial in z with complex coefficients, has at least one pair of solutions ψ_1 and φ_1, where ψ_1 and φ_1 are polynomials. This follows from Theorem V with $\delta = \frac{1}{2}$ and $F \equiv 0$ and the representation of ψ and φ in Theorem III. In fact, one may bound the degrees in z of the solutions ψ_1 and φ_1 by $\frac{1}{2}\deg r(z)$ and $\frac{1}{2}[\deg r(z)] - \frac{1}{2}$, respectively. In the case of a quadratic coefficient for φ^2 we are again guaranteed the existence of polynomial solutions, but we have no bounds on their degrees, by this argument, because of the $O(\log z)$ term in the asymptotic expansion of $\sqrt{qp}\,\gamma$.

II. METHOD AND PROCEDURE

We shall first prove the nonexistence of solutions

under the conditions asserted in Theorem I. Clearly $(z-a)^{2\alpha-1}$ divides ψ^2, so $(z-a)^{\alpha}$ divides ψ and $(z-a)^{2\alpha}$ divides both ψ^2 and g. But then $(z-a)^{2\alpha}$ divides h, contrary to hypothesis, so there are no solutions. The remainder of Theorem I is a consequence of Theorems II, III, and IV.

We shall next prove Theorem II. We begin with an example. Consider the equation

$$\psi^2 + z^2(z-1)(z-2)(z-3)^2\varphi^2 = z^2(z-1)^2(z-2)(z-4).$$

We may write this as

$$\psi^2 + [z^2(z-1)(z-2)][(z-3)^2]\varphi^2 = [z^2(z-1)^2(z-2)][z-4],$$

where the polynomials in square brackets are each relatively prime except for the first and third, which, indeed, have all their distinct linear factors in common. If the first and third factors above had no multiple zeros, we would have the desired form already. Instead, we notice that $z(z-1)$ must divide ψ and, using this, $z-1$ divides φ. Then setting $\psi = z(z-1)\psi_1$ and $\varphi = (z-1)\varphi_1$ we have

$$\psi_1^2 + [(z-1)(z-2)][(z-3)^2]\varphi_1^2 = [z-2][z-4],$$

which is in the desired form. The same sort of scheme will now be used to handle the general case.

For any two polynomials α and β let (α, β) denote their greatest common divisor. We may write (1) in the form

$$\psi^2 + Q_1Q_2\varphi^2 = Q_3Q_4 \tag{4}$$

where $(Q_1, Q_2) = (Q_3, Q_4) = (Q_3, Q_4) = (Q_2, Q_4) = 1$, i.e., let Q_1 denote the monic polynomial of least degree such that $(gQ_1^{-1}, (g,h)) = 1$, let Q_3 denote the monic polynomial of least degree such that $(hQ_3^{-1}, (g,h)) = 1$, let $Q_2 = gQ_1^{-1}$, and let $Q_4 = hQ_3^{-1}$. If Q_1 and Q_3 have no multiple zeros, the $Q_1 = Q_3$, and we are already in form (3). Suppose $z-a$ is a factor of both Q_1 and Q_3 of multiplicity m and n, respectively, where $m, n > 1$. We shall show the (4) is constructively equivalent to an equation of type (5) immediately below, i.e.,

$$\psi_1^2 + Q_{1,1}Q_{2,1}\varphi_1^2 = Q_{3,1}Q_{4,1} \tag{5}$$

where the $Q_{i,1}$ satisfy the same conditions as the corresponding Q_i, $i = 1, 2, 3, 4$, where $Q_{1,1}Q_{2,1}$ divides Q_1Q_2, where $Q_{3,1}Q_{4,1}$ divides Q_3Q_4, and where $z-a$ has a common multiplicity of either zero or one in both $Q_{1,1}$ and $Q_{3,1}$. This will suffice to prove Theorem II, since one may continue until after t iterations, for some positive integer t, $Q_{1,t}$ and $Q_{3,t}$ are identical and have no multiple zeros.

There are four possible cases. Suppose n is even and $n \leq m$. We may set $\psi = (z-a)^{n/2}\psi_1$ and cancel $(z-a)^n$ from both sides of the equation. Then set $Q_{1,1} = Q_1(z-a)^{-n}$, $Q_{2,1} = Q_2$, $Q_{3,1} = Q_3(z-a)^{-m}$, and $Q_{4,1} = Q_4(z-a)^{m-n}$. Suppose n is even and $n > m$. If m is even we may set $\psi = (z-a)^{m/2}\psi_1$, $Q_{1,1} = Q_1(z-a)^{-m}$, $Q_{2,1} = Q_2$, $Q_{3,1} = Q_3(z-a)^{-m}$, and $Q_{4,1} = Q_4(z-a)^{n-m}$. If m is odd then $(z-a)^{m+1}$ divides ψ^2 and $(z-a)^2$ divides φ^2. We may reduce to the case where m is replaced by 1 and n by $n-(m+1)$. If $n-(m+1) \geq 2$ we may continue, reducing to the cases of m replaced by 1 and n replaced by $n-(m+1)-j, j = 0, 1, \ldots, \frac{1}{2}[n-(m+1)]$ before finishing.

Suppose next that n is odd and $n < m$. This is the case where there are no solutions, by Theorem I, so we have nothing to prove. Suppose n is odd and $n \geq m$. Then $(z-a)^m$ divides ψ^2. If m is even we may set $\psi = (z-a)^{m/2}\psi_1$, $Q_{1,1} = Q_1(z-a)^{-m}$, $Q_{2,1} = Q_2$, $Q_{3,1} = Q_3(z-a)^{-m}$, and $Q_{4,1} = Q_4(z-a)^{n-m}$. If m is odd and $m = n$, then set $\psi = (z-a)^{(m-1)/2}\psi_1$, $Q_{1,1} = Q_1(z-a)^{-m+1}$, $Q_{2,1} = Q_2$, $Q_{3,1} = Q_3(z-a)^{-m+1}$, and $Q_{4,1} = Q_4$. If m and n are odd and $n > m$, then $(z-a)^{m+1}$ divides ψ^2 and $(z-a)^2$ divides φ^2. We may reduce to the cases where m is replaced by 1 and n by $n-(m+1)-2j$, $j = 0, 1, \ldots, \frac{1}{2}[n-(m+2)]$. This proves Theorem II.

III. PROOFS OF THEOREMS III–V

We shall now prove Theorem III. Suppose that ψ and φ are two entire functions which satisfy an equation of type (3). We may factor $\psi^2 + qp\varphi^2$ into $(\psi + i\sqrt{qp}\,\varphi)(\psi - i\sqrt{pq}\,\varphi)$. If we set

$$\psi + i\sqrt{qp}\,\varphi = i\sqrt{qq_1}\,\exp[i\sqrt{qp}\,\gamma(z)], \tag{6}$$

this will define, up to a choice of the branch of the logarithm, a multiple valued function $\gamma(z)$ which is certainly analytic on the complex plane with suitable cuts from the zeros of pqq_1 to $z = \infty$, as required by the statement of Theorem III. Since

$$(\psi - i\sqrt{qp}\,\varphi)(\psi - i\sqrt{qp}\,\varphi) = qq_1,$$

we see that

$$\psi - i\sqrt{qp}\,\varphi = -i\sqrt{qq_1}\,\exp[-i\sqrt{qp}\,\gamma(z)]. \tag{7}$$

If z_j is a zero of q (hence a simple zero) and if we analytically continue (6) around z_j, we obtain

$$\psi - i\sqrt{qp}\,\varphi = -i\sqrt{qq_1}\,\exp[-i\sqrt{qp}\,\hat{\gamma}(z)]$$

where $\hat{\gamma}(z)$ is the analytic continuation of $\gamma(z)$ about z_j. Then $\exp[\sqrt{qp}(i\gamma(z) - i\hat{\gamma}(z))] = 1$. Hence

$$i\sqrt{qp}\,\hat{\gamma}(z) = i\sqrt{qp}\,\gamma(z) + 2n_j\pi i,$$

for some integer n_j. Thus $\gamma(z) - n_j\pi(\sqrt{qp})^{-1}$ has an isolated singularity at $z = z_j$. Consider the identities

$$\begin{aligned} \psi(z) &= -\sqrt{qq_1}\,\sin(\sqrt{qp}\,\gamma(z)) \\ \varphi(z) &= \sqrt{p^{-1}q_1}\,\cos(\sqrt{qp}\,\gamma(z)). \end{aligned} \tag{8}$$

Suppose $z \to z_j$. If $|\sin(\sqrt{qp}\,\gamma(z))|$ has ∞ as a limit value as $z \to z_j$, then so does $|\varphi(z)|$, which is impossible for an entire function. Thus $|\sin\sqrt{qp}\,\gamma(z)|$ has only finite limit values at $z = z_j$. Let L be such a limit value. If $L \neq 0$ then $\psi(z)$ has a zero of order $\frac{1}{2}$ at $z = z_j$, which is impossible. Thus $L = 0$ and $\lim_{z \to z_j}\sqrt{qp}\,\gamma(z)$ exists and equals $n_j'\pi$ for some integer n_j'. It follows that $\sqrt{qp}(\gamma(z) - n_j\pi(\sqrt{qp})^{-1})$ has limit zero as $z \to z_j$; hence the singularity of $\gamma(z) - n_j\pi(\sqrt{qp})^{-1}$ at $z = z_j$ is removable and (i) $\gamma(z) - n_j\pi(\sqrt{qp})^{-1}$ is analytic at $z = z_j$.

Analytic continuation of $\sqrt{qp}\,\gamma(z)$ about a point z_k which is a zero of p of odd multiplicity gives us the equations

$$\begin{aligned} \sin(\sqrt{qp}\,\hat{\gamma}(z)) &= -\sin(\sqrt{qp}\,\gamma(z)), \\ \cos(\sqrt{qp}\,\hat{\gamma}(z)) &= -\cos(\sqrt{qp}\,\gamma(z)), \end{aligned} \tag{9}$$

where $\hat{\gamma}(z)$ denotes the analytic continuation of $\gamma(z)$

about $z=z_k$. Then $\hat{\gamma}(z)=\gamma(z)-(2l_k+1)\,\pi(\sqrt{qp})^{-1}$ for some integer l_k. Thus $g_k(z)=\gamma(z)+(l_k+\frac{1}{2})\,\pi(\sqrt{qp})^{-1}$ has an isolated singularity at $z=z_k$.

Since $\psi+i\sqrt{qp}\,\varphi$ is analytic in $\sqrt{z-z_k}$ at $z=z_k$ and $\psi(z_k)$, is nonzero at $z=z_k$, we see that $\log(\psi+i\sqrt{qp}\,\varphi)$, hence also $\sqrt{qp}\,\gamma(z)$, is analytic in $\sqrt{z-z_k}$ at $z=z_k$. [If the multiplicity of the zero at $z=z_k$ is even, it follows that $\sqrt{qp}\,\gamma(z)$ is analytic at $z=z_k$.] We write $\sqrt{qp}\,\gamma(z)$ as $(l_k+\frac{1}{2})\,\pi+\sqrt{qp}\,g_k(z)$ where $g_k(z)$ is singlevalued and analytic in a punctured disk centered at $z=z_k$. Since $\sqrt{qp}\,\gamma(z)$ is analytic in $\sqrt{z-z_k}$ at $z=z_k$, it follows that $g_k(z)$ has, at worst, a pole at $z=z_k$. But then, since the order of vanishing of $\sqrt{qp}$ at $z=z_k$ is not an integer, $\sqrt{qp}\,g_k(z)$ vanishes at $z=z_k$. Writing

$$\varphi=(-1)^{l_k+1}\sqrt{p^{-1}q_1}\,\sin(\sqrt{qp}\,g_k(z)),$$

we see that $\sqrt{qp}\,g_k(z)(\sqrt{p})^{-1}$ must be bounded as $z\to z_k$. Then $g_k(z)$ is analytic at $z=z_k$, if l_k is odd.

Suppose that z_k has even multiplicity. Then, from the above, $\gamma(z)$ has, at worst, a pole at $z=z_k$. Also, since

$$\varphi=\sqrt{p^{-1}q_1}\,\cos(\sqrt{qp}\,\gamma(z))$$

is entire, we must have that $\sqrt{qp}\,\gamma(z)$, which is analytic at $z=z_k$ by the comment in parenthesis above, must equal $(l_k+\frac{1}{2})\,\pi$ at $z=z_k$ for some integer l_k. Then

$$\varphi=(-1)^{l_k+1}\sqrt{p^{-1}q_1}\,\sin\{\sqrt{qp}\,[\gamma(z)-(l_k+\tfrac{1}{2})\,\pi(\sqrt{qp})^{-1}]\}.$$

Since

$$\sqrt{qp}\,[\gamma(z)-(l_k+\tfrac{1}{2})\,\pi(\sqrt{qp})^{-1}]\to 0\ \text{ as }\ |z-z_k|\to 0,$$

we see that $q(z)[\gamma(z)-(l_k+\frac{1}{2})\,\pi(\sqrt{qp})^{-1}]$ must be bounded as $z\to z_k$. Therefore, (ii) $\gamma(z)-(l_k+\frac{1}{2})\,\pi(\sqrt{qp})^{-1}$ is analytic at $z=z_k$, for all zeros z_k of $p(z)$.

Let z_r be a zero of q_1. In (6) let $\psi+i\sqrt{qp}\,\varphi$ vanish to the order $\frac{1}{2}(m_r+t_r)$ at $z=z_r$. Then $\psi-i\sqrt{qp}\,\varphi$ vanishes to the order $\frac{1}{2}(m_r-t_r)$. Also, from (6), $\sqrt{qq_1}\,\exp[i\sqrt{qp}\,\gamma(z)-\frac{1}{2}t_r\log(z-z_r)]$ is analytic and nonvanishing at $z=z_r$; hence, (iii) $\sqrt{qp}\,\gamma(z)+\frac{1}{2}t_r i\log(z-z_r)$ is analytic at $z=z_r$. Obviously t_r is an integer and so is $\frac{1}{2}(m_r+t_r)$ and $\frac{1}{2}(m_r-t_r)$. Further, $-m_r\le t_r\le m_r$.

We have proven the necessity of (i), (ii), and (iii). Formally, the functions ψ and φ given in (8) satisfy $\psi^2+qp\,\varphi^2=qq_1$. We must show that, in light of (i), (ii), and (iii), above, $\psi(z)$ and $\varphi(z)$ are each entire.

The only places where analyticity must be checked are at the zeros of qpq_1. At any z_j which is a zero of q we see that by (8) we have

$$\psi(z)=-q\sqrt{pq_1}(-1)^{n_j}(\sqrt{qp})^{-1}\sin\{\sqrt{qp}[\gamma(z)-n_j\,\pi(\sqrt{qp})^{-1}]\}),$$

which is analytic at $z=z_j$ by (i). Also

$$\varphi(z)=\sqrt{p^{-1}q_1}\,(-1)^{n_j}\cos\{\sqrt{pq}[\gamma(z)-n_j\,\pi(\sqrt{pq})^{-1}]\},$$

which is also analytic at $z=z_j$. At any point z_k which is a zero of p we have

$$\psi(z)=-\sqrt{qq_1}\,(-1)^{l_k}\cos\{\sqrt{qp}[\gamma(z)-(l_k+\tfrac{1}{2})\,\pi(\sqrt{qp})^{-1}]\}$$

and

$$\varphi(z)=\sqrt{p^{-1}q_1}\,(-1)^{l_k+1}\sin\{\sqrt{qp}[\gamma(z)-(l_k+\tfrac{1}{2}\pi)(\sqrt{qp})^{-1}]\}.$$

Thus $\psi(z)$ and $\varphi(z)$ are analytic at each point z_k by (ii).

At each point z_r, which is a zero of q_1, we have, setting $\sqrt{qp}\,\gamma(z)+\frac{1}{2}t_r i\log(z-z_r)$ equal to $\gamma_r(z)$,

$$\psi(z)=-(2i)^{-1}\sqrt{qq_1}\,((z-z_r)^{t_r/2}\exp(i\gamma_r(z))-(z-z_r)^{-t_r/2}\exp(-\gamma_r(z))]$$

and

$$\varphi(z)=\tfrac{1}{2}\sqrt{p^{-1}q_1}\,[(z-z_r)^{t_r/2}\exp(i\gamma_r(z))+(z-z_r)^{-t_r/2}\exp(-i\gamma_r(z))].$$

Since $\gamma_r(z)$ is analytic at $z=z_r$ and $(z-z_r)^\tau$ is analytic at $z=z_r$ if $\tau=\frac{1}{2}(m_r+l_r)$ or $\frac{1}{2}(m_r-l_r)$, we see that $\psi(z)$ and $\varphi(z)$ are each analytic at $z=z_r$.

To see (iv) note that $\sqrt{qp}\,\gamma$ equals $\log(\psi-i\sqrt{qp}\,\varphi)-\log(i\sqrt{qq_1})$. Thus $\sqrt{qp}\,\gamma$ is unique up to an integral multiple of $2\pi i$. This proves Theorem III.

We shall next prove Theorem IV. We shall show the existence of a polynomial $f(t)$ such that for some a in $\mathbf{C}$ which is not a zero of pqq_1, and each j, k, and r as in Theorem III,

$$(\alpha)\ \int_a^{z_j}[p(t)\,q(t)]^{1/2}[s(t)]^{-1}f(t)\,dt=n_j\pi,$$

$$(\beta)\ \int_a^{z_k}[p(t)\,q(t)]^{1/2}[s(t)]^{-1}f(t)\,dt=(l_k+\tfrac{1}{2})\,\pi,$$

and

$$(\gamma)\ \sqrt{pq}\,(s)^{-1}\text{ has a residue of }\tfrac{1}{2}t_r i\text{ at }z=z_r.$$

(To be definite let each path of integration be the straight line segment between the two point. We will choose the point a such that no two zeros of pqq_1 lie on a common line which also contains a.)

We claim that

$$\gamma_1(z)=(\sqrt{pq})^{-1}\int_a^z[p(t)\,q(t)]^{1/2}[s(t)]^{-1}f(t)\,dt$$

will satisfy (i), (ii), and (iii) of Theorem III. [Certainly $\gamma_1(z)$ will be analytic except for arbitrary cuts from each zero of qpq_1 to $z=\infty$.] At each zero z_j of q, for all $z\neq z_j$ but sufficiently close to z_j,

$$\gamma_1(z)-n_j\pi(\sqrt{pq})^{-1}=(\sqrt{pq})^{-1}\int_{z_j}^z[p(t)\,q(t)]^{1/2}[s(t)]^{-1}f(t)\,dt.$$

As one may check by expanding the integrand as a fractional power of $z-z_j$ times a Taylor series in $z-z_j$ this shows that $\gamma_1(z)-n_j\pi(\sqrt{qp})^{-1}$ is analytic at $z=z_j$. At $z=z_k$ the exact analog is true. Next we shall check the points $z=z_r$. We notice that

$$(\sqrt{qp})^{-1}\int_a^z\{[p(t)\,q(t)]^{1/2}[s(t)]^{-1}f(t)-\tfrac{1}{2}t_r i(z-z_r)^{-1}\}dt$$

is analytic at $z=z_r$, since the integrand is analytic there. Thus $\gamma(z)-\frac{1}{2}it_r\log(z-z_r)(\sqrt{qp})^{-1}$ is analytic at $z=z_r$. This proves that we need only show that (α), (β), and (γ) can be satisfied by some polynomial f.

Suppose that we have indexed the set of z_j and z_k by z_u. Then it will suffice to find a collection of polynomials f_u such that

$$\int_a^{z_u}[p(t)\,q(t)]^{1/2}f_{u_1}(t)\,dt=\delta_u^{u_1},\tag{10}$$

where $\delta_u^{u_1}$ is the Kronecker delta. Assuming (10) set

$$f=\sum_u c_u f_u+\sum_r\tfrac{1}{2}t_r i[q(z_r)\,p(z_r)]^{1/2}\,s(z)(z-z_r)^{-1},$$

where

$$c_u = -\int_a^{z_u} \sum_r \tfrac{1}{2} t_r i [q(z_r)]^{1/2} s(t)(t - z_r)^{-1}\, dt + b_u$$

and

$$b_u = n_j \pi \quad \text{if } z_u = z_j,$$

$$b_u = (l_k + \tfrac{1}{2})\pi \quad \text{if } z_u = z_k.$$

Thus (α), (β), and (γ) can be satisfied if we can satisfy (10). We wish to pick the point a so that no two of the $|z_u - a|$ are equal and so that a is not on any line containing two distinct z_u's. (These conditions can be satisfied by choosing a to be off of a finite number of lines.) Now let us change variables, sending the point a to zero and each z_u into $z_u - a \neq 0$, which we shall denote by z_u in what follows. We wish to show that the matrix

$$\left(\int_0^{z_u} [p(t)q(t)]^{1/2} t^k\, dt\right),$$

$0 \leq k < \infty$ and $1 \leq u \leq N$ (where N is the number of distinct z_u's) has rank N. If the rank is N then there exists a nonsingular matrix of the form

$$\left(\int_a^{z_u} [p(t)q(t)]^{1/2} t^{k_l}\, dt\right),$$

$1 \leq u \leq N$, and $1 \leq l \leq N$, for a set of distinct nonnegative integers k_l. Then there would exist constants d_{l,u_1}, $1 \leq u_1 \leq N$, such that each

$$\int_0^{z_u} [q(t)p(t)]^{1/2} \left(\sum_l d_{l,u_1} t^{k_l}\right) dt = \delta_u^{u_1}.$$

Setting $f_{u_1}(z) = \sum_l d_{l,u_1} z^{k_l}$ we would be through.

Suppose that the matrix above does not have rank N, then there exist complex constant $c_1, \ldots, c_N$, not all zero, such that

$$\int_0^{z_u} \sum_{u=1}^{N} c_u [q(t)p(t)]^{1/2} t^k\, dt = 0$$

for $k = 0, 1, \cdots$.

We may change variables to conclude that

$$\int_0^1 \sum_{u=1}^{N} K_u [q(tz_u)p(tz_u)]^{1/2} t^k\, dt = 0$$

for $k = 0, 1, \cdots$. This will now be shown to be impossible unless $\xi(t) \equiv \sum_{u=1}^{N} K_u [q(tz_u)p(tz_u)]^{1/2} \equiv 0$. On $[0,1]$ we may uniformly approximate the real part of $\xi(t)$ by a sequence of polynomials with real coefficients $\xi_n(t)$ by the Weierstrass approximation theorem. Then

$$0 = \lim_{n\to\infty} R\left(\int_0^1 \xi(t)\,\xi_n(t)\, dt\right)$$

$$= \lim_{n\to\infty} \int_0^1 R[\xi(t)]\,\xi_n(t)\, dt$$

$$= \int_0^1 \{R[\xi(t)]\}^2\, dt,$$

where $R(z)$ denotes for each complex z the real part of z. Thus $R[\xi(z)] \equiv 0$ on $[0,1]$. Similarly, one may show that the imaginary part vanishes identically.

We shall next prove that the $[q(tz_u)p(tz_u)]^{1/2}$ are linearly independent functions over $\mathbf{C}$. Suppose, first, that not all of the zeros of qp are of even multiplicity. Let z_{u_1} denote that zero of odd multiplicity having least absolute value. Let z_{u_2} denote, among the z_u's with $K_u \neq 0$, having largest absolute value. Then $[q(tz_{u_2})p(tz_{u_2})]^{1/2}$ has a branch point, $z_{u_1} z_{u_2}^{-1}$ which no other $[q(tz_u)p(tz_u)]^{1/2}$ possesses. Thus in this case linear dependence is impossible.

The sole remaining case is where $[q(t)p(t)]^{1/2} = v(t)$, a polynomial. Now we must see if the functions $v(lz_u)$ can be linearly dependent. Set $v(t) = \sum_{j=0}^{m} e_j t^j$ where $m \geq N$. If no e_j vanishes then our dependence relation would imply that every

$$\sum_{u=1}^{N} K_u z_u^j \equiv 0,$$

for $0 \leq j \leq N-1$. By the nonvanishing of the Vandermonde determinant this is impossible. Hence the only remaining case is when some of the e_j's vanish. Our present z_u's are of the form $z_u - a$ where the point a was chosen to be off of a finite number of lines. We wish to see that if we also choose a not to be any one of a finite number of points we could assume no e_j vanishes. Set $v(t) = \prod_{u=1}^{N} (t - z_u)^{l_u}$, for a set of positive integers l_u. Then in terms of our original z_u's we have $v(t) = \prod_{u=1}^{N} (t + a - z_u)^{l_u}$, choosing the point a such that $v(0)\, v^{(1)}(0) \cdots v^{(\gamma)}(0) \neq 0$, where $\gamma = \sum_{u=1}^{N} l_u$, we may assume, above, that no $e_j = 0$, This proves Theorem IV.

Proof of Theorem V: We use the formula

$$\sqrt{p(z)q(z)}\ \gamma(z) = \left(\int_a^z \sqrt{p(t)q(t)}\,[s(t)]^{-1} f(t)\, dt\right) + (\sqrt{pq})\, g,$$

where g is an entire function. Now

$$\sqrt{p(t)q(t)}\,[s(t)]^{-1} f(t)$$

has an expansion about $z = \infty$ in descending integral powers of t, if δ is an integer, or descending odd integral powers of $t^{1/2}$, if δ is not an integer. Thus the integral is of the same form as the integrand, except that if δ is an integer there may now be a log$z$ term, and if δ is not an integer, there may now be a constant term. If δ is an integer set the terms where the power of z is nonnegative equal to G. If δ is not an integer set the terms where the power of z is positive equal to G. We may find a series h, in descending integral powers of z, such that

$$h = (\sqrt{pq})^{-1}(F - G),$$

if δ is an integer, or

$$h = (\sqrt{pq})^{-1}(\sqrt{z}\, F - G),$$

if δ is not an integer. Let the polynomial g be defined to be such that $h - g$ vanishes at $z = \infty$. Then

$$\sqrt{pq}\ g = (F - G) + O(z^{\delta-1}),$$

if δ is an integer, and

$$\sqrt{pq}\ g = (\sqrt{z}\, F - G) + O(z^{\delta-1})$$

if δ is not an integer. With this choice of g we obtain the desired asymptotic form for $\sqrt{pq}\ \gamma$, up to an error which varies according to the cases. If δ is an integer,

the error is the larger of $O(z^{\delta-1})$ and $O(\log z)$. If δ is not an integer the error is the larger of $O(z^{\delta-1})$ and $O(1)$. This proves parts (i), (ii), and (iii) of Theorem V. We next prove part (iv).

By the remark after Theorem IV we see that we must have

$$\sqrt{pq}\,\gamma_1 - \sqrt{pq}\,\gamma_2 = \sqrt{pq}\,g$$

where g is entire. Also $\sqrt{pq}\,g$ is asymptotic to $O(z^{\delta-1})$, $O(\log z)$, or $O(1)$, as the case may be. In the first and thirds cases, since $\delta > 0$,

$$\lim_{n\to\infty} |g(z)| = 0, \quad \text{so } g \equiv 0.$$

Also, in the second case, if $\delta = 1$, the same conclusion holds. This proves Theorem V.

[1]J. L. Lebowitz and O. Penrose, J. Math. Phys. **13**, 604 (1972).

[2]Consider the function $E(z)$ given in W. K. Hayman, *Meromorphic Functions* (Oxford U. P., Oxford, 1964), p. 81, and set $E_1(z) = E(z + Ki)$ for any real K with $|K| > \pi$.

A superposition principle for Siegert resonant states*

David L. Huestis†

Arthus Amos Noyes Laboratory of Chemical Physics,‡ California Institute of Technology, Pasadena, California 91125
(Received 5 May 1975)

The utilization of the Siegert resonant states as a natural basis for the scattering process has been constrained by the lack of a convergent superposition formula. This paper presents such an expansion of the exact scattering wavefunction over the *s*-wave Siegert resonant states for finite range potentials. Any integral property of the scattering wavefunction can then be calculated from the resonances.

INTRODUCTION

Since the early days of the quantum theory of scattering, metastable states have served as an appealing explanation for the structure observed in cross sections. In the discussion of some specific physical interaction, say $\pi+p$ or e^-+H_2, these compound states of the target and projectile are associated with a pole of the S matrix. Breit and Wigner[1] gave a simple model formula that describes the cross section near such a resonance:

$$\sigma \simeq \frac{\pi}{k^2}\,\frac{\Gamma^2}{(E-E_{\rm res})^2+\frac{1}{4}\Gamma^2} \tag{1}$$

where Γ is the width of the resonance and $2/\Gamma$ is the lifetime of the resonant state. This important one-level formula is a satisfactory description when the widths are narrow compared with the energy separation between the resonances. In the more general case one expects to use a superposition of Breit–Wigner terms with relative phases chosen in some yet unspecified manner.

Regge[2] provided a remarkably simple albeit slowly convergent[3] expression for the S matrix by a Weierstrass–Hadamard infinite product over all the resonances

$$S(k)=\exp(-2ikr_0)\prod_n \frac{(k_n+k)}{(k_n-k)}, \tag{2}$$

where r_0 is the actual radius of the potential. Regge's formula remains correct regardless of where the resonances are located, provided the potential is of finite range. This formula gives an expansion over the resonances from the purely mathematical properties of the S matrix, but it camouflages the physically appealing role of the resonances as transient intermediates for the scattering process.

The objective here is more than just to describe the shape of cross section or S matrix in terms of dispersion formulae such as (1) and (2). Rather, the basic ideas are: (i) the resonances are compound states, (ii) the incoming particle has some amplitude to populate each of these states, and (iii) it is the subsequent decay of the resonant states that is observed

$$\psi_{\rm in} \rightarrow \sum_n \alpha_n\varphi_n + \psi_{\rm out} \tag{3}$$

incoming → resonances + unscattered.

↓

decay

The problem at hand is to characterize the resonance energies and resonant states as manifestations of the potential, then to describe the scattering process as mediated by the resonances.[4]

In particular, we need to develop some kind of superposition principle for resonant states:

$$\psi_{\rm given}=\sum_n \alpha_n\varphi_n. \tag{4}$$

This would presumably illustrate that once a given wavefunction is expanded, any property could be written in terms of the simpler and decoupled properties of the resonant states themselves:

$$\text{property}[\psi_{\rm given}] \leftarrow \text{properties}[\varphi_n]. \tag{5}$$

Such a superposition principle we now set out to find.

THE RESONANT STATES

Kapur and Peierls[5] suggested that the eigenvalue problem

$$-\frac{\hbar^2}{2m}\theta_n''(r)+V(r)\theta_n(r)=E_n\theta_n(r), \quad 0\le r\le r_0, \tag{6}$$

$$\theta_n(0)=0, \quad \theta_n'(r_0)=ik\theta_n(r_0)$$

for a potential of finite range r_0 would provide the s-wave resonance energies as eigenvalues $E_n=E_{\rm res}+(i/2)\Gamma$. It is crucial to note that each solution behaves as an outgoing plane wave at the right-hand boundary, but that k is chosen to match the momentum of the incoming particle we wish to describe. The resulting eigenfunctions $[\theta_n]$ are called the Kapur–Peierls states. A superposition formula was derived from these states for the exact scattering wavefunction

$$\psi_k \rightarrow \sum_n \beta_n(k)\,\theta_n. \tag{7a}$$

Whence a dispersion formula for the transition matrix was available:

$$T=\frac{\hbar^2}{2m}\exp(-2ikr_0)\sum_n \frac{\theta_n^2(r_0)}{(E_n-E)N_n}-\frac{1}{k}\sin kr_0 \times\exp(-ikr_0), \tag{7b}$$

where N_n is the normalization integral $N_n=\int_0^{r_0}\theta_n^2(r)\,dr$.

This model has attracted attention for two reasons. First, it appears to offer a technique for finding the resonances for a variety of potentials,[6] even for problems with many particles.[7] Second, the expansions (7a) and (7b) over a complete[8] set of resonant states appears to provide a description of all of the properties and structures of the scattering process in terms of the

resonances. Unfortunately there is one principal difficulty. The Kapur—Peierls formula (7b) suggests that the S matrix should have poles at the Kapur—Peierls energies. Although these energies may in some cases be close to the actual poles, they are not coincident. Indeed the poles of the S matrix are characteristic of the potential itself, whereas the Kapur—Peierls energies depend on the incoming momentum k, and in addition upon the choice of r_0. This is true even when r_0 is chosen outside the actual range of the potential!

Siegert[9] noticed this difficulty and suggested an alternative eigenvalue problem:

$$-\frac{\hbar^2}{2m}\varphi_n'' + V\varphi_n = E_n\varphi_n, \quad \frac{\hbar^2}{2m}k_n^2 = E_n,$$
$$\varphi_n(0)=0, \quad \varphi_n'(r_0) = ik_n\varphi_n(r_0). \tag{8}$$

We call this the Siegert eigenvalue problem; the eigenvalues are called the Siegert energies or Siegert poles; and the eigenfunctions are called the Siegert states. Use of the resonance momentum in the plane-wave boundary condition alleviates some of the unsatisfactory features of the Kapur—Peierls description. In particular one can easily show[9] (i) that the Siegert energies are at the poles of the S matrix, (ii) that they are independent of both the actual scattering energy and the joining radius (as long as it is chosen outside the actual potential), and finally (iii) that the bound states are just special solutions of the Siegert equations (8).

Siegert also found a one-level formula that gives the correct residue of the S matrix at the resonance pole and reduces to the Breit—Wigner formula when the width Γ is small. However, he was unable to provide a complete description of the scattering process due to the lack of an expansion over the resonant states. Any attempt to remedy this situation is complicated by the nonorthogonality of the Siegert states (which arises from having the boundary condition depend on the eigenvalue), and by the rather slow convergence of the residues. These are the difficulties to be overcome below.

THE TWO-COMPONENT WAVEFUNCTION

We first rewrite the Siegert s-wave eigenvalue problem in a more dimensionless form:

$$-\varphi_n'' + U\varphi_n = k_n^2\varphi_n, \quad \varphi_n(0)=0, \quad \varphi_n'(r_0) = ik_n\varphi_n(r_0),$$
$$k_n^2 = (2m/\hbar^2)E_n, \quad U=(2m/\hbar^2)V. \tag{9}$$

Unfortunately, the literature of such problems with the eigenvalue in the boundary conditions is meager. However, Friedman[10] discusses two similar problems:

$$-\varphi'' = k^2\varphi, \quad \varphi(0)=0, \quad \varphi'(1)=k^2\varphi(1),$$
and
$$-\varphi'' = k^2\varphi, \quad \varphi(0)=0, \quad \varphi'(1)=k\varphi(1). \tag{10}$$

A further description of these two and other related systems can be found in Ref. 3.

An important suggestion of the work in Refs. 3 and 10 is that we consider writing the solutions to Eq. (9) as two-component wavefunctions

$$\Phi_n = \begin{bmatrix} \varphi_n^{(1)}(r) \\ \varphi_n^{(2)}(r) \end{bmatrix}. \tag{11}$$

Then the eigenvalue problem is rewritten as

$$\mathbf{L}\Phi_n = \begin{bmatrix} \varphi_n^{(2)\prime}(r) \\ \varphi_n^{(1)\prime}(r) + \int_r^{r_0} U(x)\varphi_n^{(1)}(x)\,dx \end{bmatrix} = ik_n\Phi_n, \tag{12a}$$

with the boundary conditions

$$\varphi_n^{(n)}(0)=0, \quad \varphi_n^{(1)}(r_0)=\varphi_n^{(2)}(r_0). \tag{12b}$$

The upper component $\varphi_n^{(1)}(r)$ from a solution Φ_n of Eq. (12a) can be seen to be a solution to (9) by differentiation of the lower components of (12a)

$$\varphi_n^{(1)\prime\prime} - U\varphi_n^{(1)} = ik_n\varphi_n^{(2)\prime} = -k_n^2\varphi_n^{(1)}. \tag{13}$$

In this two-component space we define the inner product by

$$\langle \Psi, \Phi\rangle = \int_0^{r_0}(\psi^{(1)}\varphi^{(1)} - \psi^{(2)}\varphi^{(2)}) \tag{14}$$

where

$$\Psi = \begin{bmatrix} \psi^{(1)} \\ \psi^{(2)} \end{bmatrix}, \quad \Phi = \begin{bmatrix} \varphi^{(1)} \\ \varphi^{(2)} \end{bmatrix}.$$

With the boundary conditions (12b) and the inner product (14) we can find the adjoint operator $\mathbf{L}^\dagger$:

$$\mathbf{L}^\dagger\Phi = \begin{bmatrix} \varphi^{(2)\prime}(r) - U(r)\int_0^r \varphi^{(2)}(x)\,dx \\ \varphi^{(1)\prime} \end{bmatrix}. \tag{15}$$

This defines a new set of adjoint eigenvectors (with the old eigenvalues)

$$\mathbf{L}^\dagger\Phi_n^\dagger = ik_n\Phi_n^\dagger. \tag{16}$$

We notice further that if the boundary conditions (12b) are satisfied, then

$$\mathbf{L}^2\Phi_n = \begin{bmatrix} \varphi_n^{(1)\prime\prime} - U\varphi_n^{(1)} \\ \varphi_n^{(2)\prime\prime} + \int_r^{r_0} U(x)\varphi_n^{(2)\prime}(x)\,dx \end{bmatrix} = -k_n^2\Phi_n,$$
$$(\mathbf{L}^\dagger)^2\Phi_n^\dagger = \begin{bmatrix} \varphi_n^{\dagger(1)\prime\prime} - U\varphi_n^{\dagger(1)} \\ \varphi_n^{\dagger(2)\prime\prime} - \{U(r)\int_0^r \varphi_n^{\dagger(2)}\}' \end{bmatrix} = -k_n^2\Phi_n^\dagger. \tag{17}$$

Equations (12), (15), and (17) show that the upper components of $\{\Phi_n\}$ and $\{\Phi_n^\dagger\}$ can be chosen to be the same and that this function, $\varphi_n = \varphi_n^{(1)} = \varphi_n^{\dagger(1)}$, is the physical Siegert resonant wavefunction. The lower components are yet without physical interpretation.

Toward the goal of a superposition principle, it will be necessary to evaluate the inner product of $\Phi_n^\dagger$ with Φ_l. As might be expected the choice of the form of the inner product and the relation of $\mathbf{L}^\dagger$ to $\mathbf{L}$ make this especially simple:

$$i(k_n - k_l)\langle\Phi_n^\dagger, \Phi_l\rangle = \langle \mathbf{L}^\dagger\Phi_n^\dagger, \Phi_l\rangle - \langle\Phi_n^\dagger, \mathbf{L}\Phi_l\rangle = 0, \tag{18}$$

so that either $\langle\Phi_n^\dagger, \Phi_l\rangle = 0$, or $k_l = k_n$. The solutions are in general nondegenerate,[4a] so we have

$$\langle\Phi_n^\dagger, \Phi_l\rangle = N_l\delta_{nl}, \tag{19}$$

where δ_{nl} is the Kronecker delta and N_l may be considered the (not necessarily real) norm of the state. As is customary the normalization constant could be absorbed into the wavefunction, yielding $N_l = 1$, except that a Siegert state could possibly have a norm of zero. Indeed we anticipate null vectors and degenerate roots to appear simultaneously[11] but only for special values of the potential strength.

Now we have our superposition principle. To expand a given two-component vector

$$\mathbf{F} = \begin{bmatrix} f^{(1)} \\ f^{(2)} \end{bmatrix} \to \sum_n \alpha_n \Phi_n, \tag{20}$$

we only need to calculate

$$\alpha_l = \frac{1}{N_l}\langle \Phi_l^\dagger, \mathbf{F}\rangle = \frac{1}{N_l}\int_0^{r_0} (\varphi_l^{\dagger(1)} f^{(1)} - \varphi_l^{\dagger(2)} f^{(2)}). \tag{21}$$

EXPANSION OF THE PHYSICAL SCATTERING STATE

For any energy there exists a function ψ_k satisfying

$$-\psi_k'' + U\psi_k = k^2\psi_k, \quad \psi_k(0) = 0. \tag{22}$$

This (ψ_k) is the exact scattering wavefunction. The principal goal of this work is to find a set of coefficients $\alpha_l(k)$ such that

$$\psi_k(r) = \sum_n \alpha_n(k)\,\varphi_n(r); \tag{23}$$

that is, to express ψ_k as a superposition of Siegert resonances.

The function ψ_k specifies only the upper component of the vector to be expanded:

$$\Psi_k = \begin{bmatrix} \psi_k \\ ? \end{bmatrix} = \sum_n \alpha_n(k)\,\Phi_n. \tag{24}$$

At this point it does not seem too important what lower component is used. Physically, we are only interested in the upper component, at least for now. There are two difficulties. First, even though the upper component of Ψ_k seems independent of the lower one, the expansion coefficients will depend on both components through Eq. (21). This is the same as saying that there are many sequences of coefficients $\{\alpha_n\}$ that give the same expanded upper component, or as noted by More and Gerjuoy,[13] the Siegert basis is overcomplete. The series (23) will converge very slowly or not at all if the lower component is not chosen with care. Second, the expansion (23) will be essentially useless if we actually have to evaluate integrals involving the yet unknown ψ_k. If the Siegert states are in fact a natural basis for the scattering problem, the stationary wavefunctions, ψ_k would be expressable directly. That is what Kapur and Peierls tried to do.

A little experimentation with Fourier series convinces one that the attempt to expand a function that does not satisfy the same boundary conditions as the basis set produces *slow* and *nonuniform* convergence (see also Ref. 3). The problem is at its worst when expanding in Siegert states. The Siegert states, even when normalized to unity are not bounded with respect to k_n. The point in the interval at which convergence will be the most difficult to obtain will be $r = r_0$, since there the Siegert states will in general by the largest $(|\varphi_n(r_0)| \sim \exp[|Im(k_n)|r_0])$.

It was found[3] that an expansion over Siegert states [Eq. (20)] will not converge at all unless at least three conditions are met:

$$\begin{aligned} &(1)\ f^{(1)}(0) = 0, \\ &(2)\ f^{(1)}(r_0) = f^{(2)}(r_0), \\ &(3)\ f^{(1)\prime}(r_0) = f^{(2)\prime}(r_0). \end{aligned} \tag{25}$$

These conditions amount to making the vector to be expanded as continuous as possible at the right-hand boundary. For the problem at hand, the lower component of Ψ_k is uniquely defined if Ψ_k is required to satisfy both (25) and the very useful relation

$$\mathbf{L}^2\Psi_k = -k^2\Psi_k. \tag{26}$$

This equation is a relatively natural choice since its upper component is identical to Eq. (22).

All of this gives

$$\Psi_k = \begin{bmatrix} \psi_k \\ a\int^r \psi_k + b\int^r \bar{\psi}_k + c \end{bmatrix} \tag{27}$$

where $\bar{\psi}_k$ satisfies the conditions

$$-\bar{\psi}_k'' + U\bar{\psi}_k = k^2\bar{\psi}_k, \quad \bar{\psi}'(0) = 0, \tag{28}$$

and a, b, and c are constants to be determined from (25) and (26). Having chosen Ψ_k we need only evaluate

$$\alpha_n(k) = \langle \Phi_n^\dagger, \Psi_k\rangle / N_n. \tag{29}$$

To evaluate α_n we use a trick much like that used by Kapur and Peierls: From (16) we have

$$\Phi_n^\dagger = \frac{\mathbf{L}^{\dagger 2} + k^2}{(ik_n)^2 + k^2}\Phi_n^\dagger, \tag{30}$$

so that

$$\alpha_n = \frac{1}{(k^2 - k_n^2)N_n}\{k^2\langle \Phi_n^\dagger, \Psi_k\rangle + \langle \mathbf{L}^{\dagger 2}\Phi_n^\dagger, \Psi_k\rangle\}. \tag{31}$$

But now

$$\begin{aligned} \langle \mathbf{L}^{\dagger 2}\Phi_n^\dagger, \Psi_k\rangle &= \langle \mathbf{L}^\dagger\Phi_n^\dagger, \mathbf{L}\Psi_k\rangle - [[\mathbf{L}^\dagger\Phi_n^\dagger]^{(1)}\psi_k^{(2)} - [\mathbf{L}^\dagger\Phi_n^\dagger]^{(2)}\psi_k^{(1)}]_0^{r_0} \\ &= \langle \mathbf{L}^\dagger\Phi_n^\dagger, \mathbf{L}\Psi_k\rangle + 0. \end{aligned} \tag{32a}$$

Next

$$\begin{aligned} \langle \mathbf{L}^\dagger\Phi_n^\dagger, \mathbf{L}\Psi_k\rangle &= \langle \Phi_n^\dagger, \mathbf{L}^2\Psi_k\rangle - [\varphi_n^{\dagger(1)}[\mathbf{L}\Psi_k]^{(2)} - \varphi_n^{\dagger(2)}[\mathbf{L}\Psi_k]^{(1)}]_0^{r_0} \\ &= \langle \Phi_n^\dagger, \mathbf{L}^2\Psi_k\rangle - b\bar{\psi}_k(0)\varphi_n^{\dagger(2)}(0). \end{aligned} \tag{32b}$$

Finally, using (31) and (26)

$$\alpha_n(k) = \frac{-b\bar{\psi}_k(0)\varphi_n^{\dagger(2)}(0)}{(k^2 - k_n^2)N_n} = \frac{-b\bar{\psi}_k(0)\varphi_n'(0)}{(k^2 - k_n^2)N_n(-ik_n)}. \tag{33}$$

The term $-b\bar{\psi}_k(0)$ is unknown until we have solved for the exact scattering wavefunction. But it is independent of n, so it will be replaced by $A(k)$ and treated as a normalization constant to be determined later or ignored altogether. Thus the expansion formula becomes

$$\psi_k(r) = A(k)\sum_n \frac{1}{(k^2 - k_n^2)}\frac{\varphi_n'(0)}{(-ik_n)N_n}\varphi_n(r). \tag{34}$$

This is our superposition formula. It specifies how

TABLE I. Some Siegert s-wave eigenvalues for the finite square well, depth $U=-\frac{1}{2}$, radius $r_0=1$.

n[a]	Rek_n	Imk_n	Rek'_n	Imk'_n
0	0.0	−1.6506	0.0	−1.4914
1	4.0993	−2.6555	4.1423	−2.6279
2	7.4268	−3.1404	7.4553	−3.1284
3	10.661	−3.4642	10.682	−3.4573
4	13.859	−3.7083	13.876	−3.7038
5	17.040	−3.9044	17.054	−3.9012
6	20.210	−4.0684	20.222	−4.0660
7	23.373	−4.2094	23.384	−4.2075

[a] The corresponding resonances with $n<0$ satisfy $k_n=-k_{-n}^*$.

the exact stationary state ψ_k is to be expanded over the Siegert resonant states φ_n. Aside from the trivial normalization factor $A(k)$, the coefficients have a simple dependence on k,

$$\alpha_n(k)\sim\frac{1}{k^2-k_n^2}. \tag{35}$$

The only significant property of $A(k)$ is that $A(k)\sim(k^2-k_l^2)$ as $k\to k_l$, so that $\psi_k\to\varphi_l$ as $k\to k_l$ just as we would want.

We note as well that the nonphysical lower components of both Ψ_k and Φ_n have disappeared from the formula. Having this principle of superposition, we can now proceed to calculate the exact wavefunction at any energy from Siegert states. The various properties of the wavefunction can also be calculated, such as the phase shift for which we would use the Lippmann–Schwinger equation to avoid convergence difficulties in taking derivatives.

THE SQUARE WELL

The above description of the scattering process is not an attempt to give yet another dispersion formula. It was hoped that the Siegert states would provide a practical, discrete (to a good approximation finite), and calculable basis with which to describe scattering events. Along this line it is necessary to consider an example. In fact, the example has been the fulcrum of the theory.

Specifically we consider

$$U(r)=\begin{cases}-\frac{1}{2}, & 0\le r\le r_0=1\\ 0, & r>r_0\end{cases}. \tag{36}$$

This is a rather weak potential; the resonances are far from the real axis; and therefore the theory should be hard pressed. The Siegert states are

$$\Phi_n=\begin{bmatrix}\sin k'_n r\\ -i\frac{k_n}{k'_n}\cos k'_n r+i\left(\frac{k_n}{k'_n}-\frac{k'_n}{k_n}\right)\cos k'_n\end{bmatrix}, \tag{37a}$$

$$\Phi_n^\dagger=\begin{bmatrix}\sin k'_n r\\ -i\frac{k'_n}{k_n}\cos k'_n r\end{bmatrix} \tag{37b}$$

where $k_n'^2=k_n^2-U=k_n^2+\frac{1}{2}$. The resonance eigenvalues satisfy [from Eq. (12b)]

$$k'_n\cos k'_n=ik_n\sin k'_n, \tag{38}$$

which can be solved by Newton's method or as in Ref. 4a. A few of the calculated Siegert resonance eigenvalues are given in Table I.

To accomplish the expansion as in Eq. (34), we will need to know the norm, $N_n=\langle\Phi_n^\dagger,\Phi_n\rangle=1+i/k_n$, and $\varphi'_n(0)=k'_n$. So then

$$\psi_k(r)=A(k)\sum_n\frac{1}{k^2-k_n^2}\,\frac{k'_n}{(-ik_n)(1+i/k_n)}\sin k'_n r. \tag{39}$$

Comparisons of partial expansions of this uniformly convergent series[14] with the known wavefunctions are shown in Figs. 1 and 2. The appropriate value of $A(k)$ was separately calculated in each case.

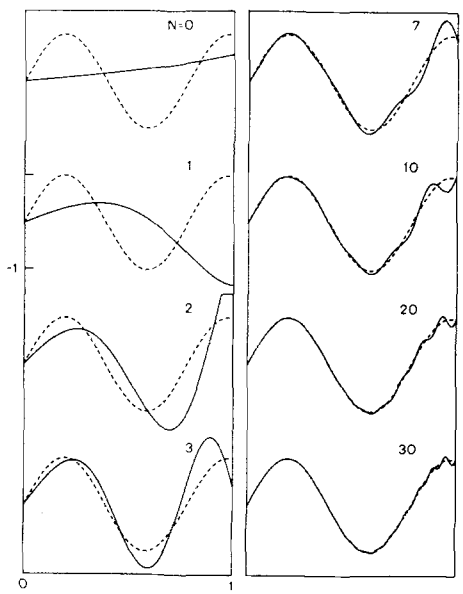


FIG. 1. Expansion $\psi_k(r)=\sum_{-N}^{N}\alpha_n\varphi_n(r)$ over square well Siegert states for $k=8$. Dashed line indicates the exact solution; the solid lines represent the partial expansion to $\pm N$. The imaginary parts of the partial expansions cancel when the terms $\alpha_n\varphi_n$ and $\alpha_{-n}\varphi_{-n}$ are included together.

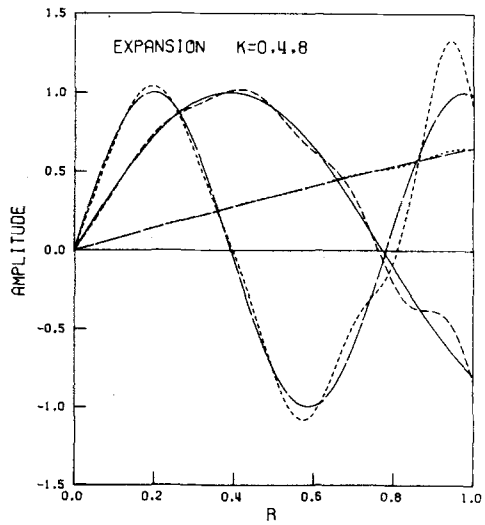


FIG. 2. Square well Siegert expansion $\psi_k(r)=\sum_{-6}^{6}\alpha_n\varphi_n(r)$ for $k=0$ (medium dashes), $k=4$ (long dashes), and $k=8$ (solid). Shorter dashed lines represent the partial expansions.

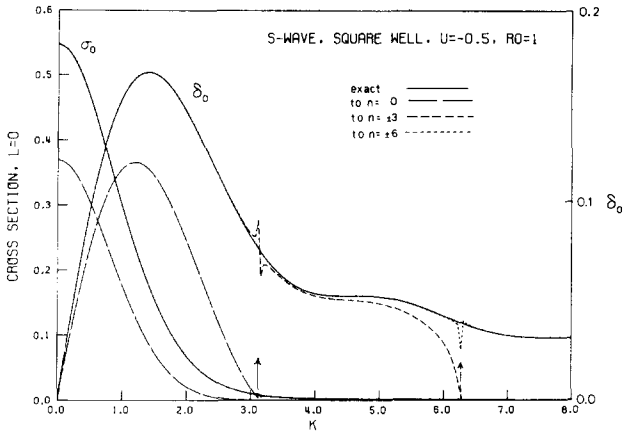


FIG. 3. Siegert expansion of δ_0 and σ_0. The singular behavior near $k=\pi$, 2π is due to the division by $\sin(kr_0)$ in the Lippmann—Schwinger evaluation of the phase shift. As indicated, the solid lines show the exact values, the dashed lines show partial expansions.

As a final illustration, we have in Fig. 3 plots of the phase shift δ_0, calculated from the Lippmann—Schwinger equation[14] along with the cross section $\sigma_0 = (\pi/k^2) \times \sin^2\delta_0$, for various numbers of exapnsion terms. The exact result is shown for comparison. We note the excellent agreement even with only a few terms in the expansion.

CONCLUSION

This paper has discussed an expansion of the s-wave scattering wavefunction over the Siegert resonant states, which is uniformly convergent at least for the finite square well. This is not to be viewed as the derivation of a dispersion formula but rather as the beginning of a calculational device. If the Siegert states turn out to resemble the bound states as much as it now appears that they do, then all of the powerful and successful approximation techniques that have been used in bound state quantum mechanics may then be used for scattering states. Specifically in view is the calculation of the Siegert states for a molecular system, such as $N_2 + e$, by some kind of self-consistent field procedure. The scattering problem is then just a rigorous expansion over the boundlike Siegert states. No fictitious "potential scattering" term or matching at the boundary would be required.

The sense of accomplishment that might be expected must be deflated by one important observation. All of the results presented apply, as yet, only to potentials of finite range (or at best dying faster than any exponential). Few physical problems have potentials of such short range. But resonances are observed nevertheless. The experience gained in this research indicates (to the author at least) that the work now in progress on Siegert states for long-range potentials will bring similar results. At that point the theory will have come to the motivation: electron—molecule scattering.

ACKNOWLEDGMENTS

It is my pleasure to acknowledge the helpful suggestions of Professor J.K. Knowles and Professor T.C. McGill. In particular I would like to thank Professor W.A. Goddard III for his unfailing interest, encouragement, enthusiasm, and support, not to mention many discussions and suggestions.

*Partially supported by National Science Foundation (Grant GP-15423) and by the President's Fund Grant (PF-013) of the California Institute of Technology. Taken in part from a thesis submitted 19 September 1972 to the Division of Chemistry and Chemical Engineering in partial fulfillment of the requirements for the Ph.D. degree, California Institute of Technology, 1973.

†National Science Foundation Predoctoral Fellow, 1968-72. Present address: Molecular Physics Center, Stanford Research Institute, Menlo Park, California 94025.

‡Contribution No. 5085

[1]G. Breit and E.P. Wigner, Phys. Rev. **49**, 519 (1936).
[2]T. Regge, Nuovo Cimento **8**, 671 (1958).
[3]D.L. Huestis, Ph.D. thesis, California Institute of Technology (1973). Available from University Microfilms.
[4]The emphasis in this work will be on the wavefunctions for the resonant states. H.M. Nussenzvieg [(a) Nucl. Phys. **11**, 499 (1959) and (b) Nuovo Cimento **20**, 694 (1961); "Analytic Properties of Nonrelativistic Scattering Amplitudes," Escuela Latino Americana de Fisica, Universidad de México, Mexico, D.F. (1962)] and G. Beck and H.M. Nussenzvieg [(c) Nuovo Cimento **16**, 416 (1960)] also capitalized on the resonances as transient intermediates in their development of resonance propagators. See also Ref. 12.
[5]P.L. Kapur and R.E. Peierls, Proc. Roy. Soc. (Lond.) A **166**, 277 (1938).
[6]D.T. Birtwistle and A. Herzenberg, J. Phys. B **4**, 53 (1971).
[7]A. Herzenberg and F. Mandl, Proc. Roy. Soc. (Lond.) A **274**, 253 (1963).
[8]R.E. Peierls, Proc. Camb. Phil. Soc. **44**, 242 (1948).
[9]A.J.F. Siegert, Phys. Rev. **56**, 750 (1939).
[10]B. Fiedman, *Principles and Techniques of Applied Mathematics* (Wiley, New York, 1956), pp. 205—207.
[11]See P.N. Dobson, Jr., J. Math. Phys. **12**, 1207 (1971), and Ref. 10.
[12]R.M. More and E. Gerjuoy, Phys. Rev. A **7**, 1288 (1973).
[13]The series is majorized by $n^{-3/2}$.
[14]T.-Y. Wu and T. Ohmura, *Quantum Theory of Scattering* (Prentice-Hall, Englewood Cliffs, N.J., 1962), p. 64.

Surface scattering on topological networks*

Aaron B. Budgor†

Institute for Fundamental Studies, Departments of Chemistry and Physics, University of Rochester, Rochester, New York 14627
(Received 12 May 1975)

The quantum mechanical network model of electrons in solids as introduced by Montroll is used to construct a multiple-scattering formalism for scattering in slightly aperiodic lattices. We examine both the case of Bloch electron scattering and diffractive scattering (as applied to LEED) from surfaces. In the former case the resulting equations are completely analogous to those derived by Koster and Maradudin in their analyses of the scattering of electron and lattice waves by defects. In the latter case the differential cross section functionally resembles Fraunhofer diffraction in optical systems, and is explicitly dependent upon the atomic geometry and the atomic potentials of the surface. All computations are analytic. Since qualitative agreement with experiment is excellent, our results seem to imply that the geometric structure of a material dominates the electronic character of a scattering system.

INTRODUCTION

The theory of scattering has been employed with much success in the investigation of the nature of matter. As first developed by Lord Rayleigh to explain the color of the sky, it has since been used as an intricate probe into the structure of liquids and solids. Up until recently there have been few experimental procedures available to study the electronic, vibronic, and geometric properties of surfaces since it requires quite a delicate interplay between the magnitude of the particle-surface interaction and the energy (and/or type) of the incident particle.[1] If the strength of the interaction is too weak, as it is for x rays, the incident particle may only be used to measure the bulk properties. If the interaction is too strong and the incident particle has a large mass, or is highly energetic, then the surface is prone to damage.

Two techniques that have found widespread applicability are elastic low energy electron diffraction (ELEED) and inelastic low energy electron diffraction (ILEED). ELEED is used as a probe of atomic geometry and the surface vibrational properties of macroscopic crystalline materials, whereas ILEED measures the dispersion relations for surface excitation spectra.

The first LEED experiment was reported by Davisson and Germer in 1927.[2] In recent years the physics of this scattering process seems to be well established. As an incident particle approaches the lattice it induces a surface charge resulting in the emission of surface plasmons. This inelastic electron–valence electron interaction contributes to forward scattering with the diffractive phenomenon caused by electron–ion core scattering. Thus, as most of the electron scattering occurs within a distance of one half the inelastic electron mean free path from the surface, that is, from 1 to 10 layers from the surface, this results in the total scattered intensity being modulated via the inelastic processes. The two most useful incident particle energy ranges for surface crystallography occur for 25 eV $\leq E \leq$ 500 eV. ELEED is most surface sensitive in the former energy region. Since the elastic and inelastic cross sections are comparable in magnitude, the intensity profiles appear highly dynamical in character with complicated secondary peaks (fine structure) occurring at all energies, but most pronounced near the Bragg angles. This behavior is indicated schematically in Fig. 1.

In this paper we present a network formalism based upon a model of electrons in solids as introduced by Montroll and co-workers[3,4,5] to describe the interaction of an assembly of scatterers in a nearly periodic material with a wave. The incident scattering particle may be of any type, i.e., electrons and x rays, as long as it interacts with the electron clouds of the scatterer. The scattering is postulated to take place on a topological network of atoms or molecules in such a way that the scattered wave may only propagate along the bonds of the network, that is, along those lines connecting atoms, nodes, together. This concept gains credence from the work of Keller who by using the multiple scattering formalism developed by Lloyd showed that above muffin tin zero the greatest angular cross sections are along the bonds of a solid, be it crystalline or amorphous.[6]

We begin in Sec. 2 with a short description of the quantum mechanical network model of electrons in solids. Section 3 contains a discussion on the use of the lattice Green's function in obtaining the band structure of a perfect solid with a small number of defects. For

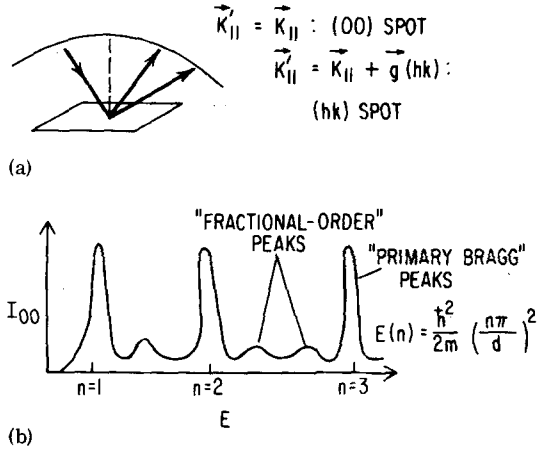


FIG. 1. (a) Intensity pattern: Fixed incident beam energy. (b) Intensity profile: Energy dependence of the intensity of a given spot.

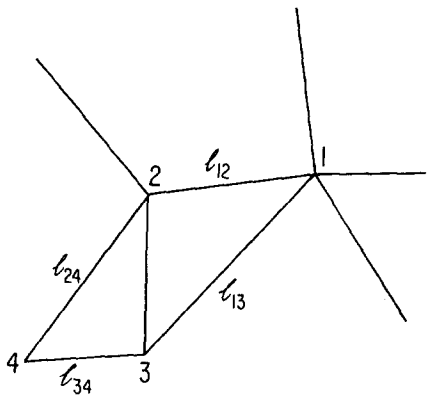


FIG. 2. An arbitrary network of atoms, $1,2,\ldots,i,\ldots,j,\cdots$, connected by bonds b_{ij} of length l_{ij}.

the purpose of the subsequent discussion the concept of a surface defect will be introduced. In Sec. 4 a formalism is constructed for the multiple scattering of a wave off of a finite number of defects in an otherwise perfect crystal and is akin to the Lippman–Schwinger[7] development of scattering theory. No Born approximation is involved. The resulting scattering equations are completely analogous to those derived by Koster and Maradudin in their analyses of the scattering of electron and lattice waves by defects.[8,9] The total wavefunction after the scattering phenomenon is made up of an incident and scattered wavefunction, the latter containing as factors the interaction of the incident wave with the defect and its subsequent propagation to other points in the solid. This Green's function propagator will be shown to asymptotically resemble an outgoing wave whose character is determined by the group velocity of the electron. As an example toward the use of this procedure a sample T-matrix calculation for a 1D and 3D s.c.c. lattice with a surface defect will be performed in Sec. 5. Finally, Secs. 6 and 7 will demonstrate how the lattice Green's function may be employed in the study of the surface structure of solids.

2. BASIC NETWORK EQUATIONS AND BAND STRUCTURE

Recently a network model of electrons in solids has been introduced by Montroll and his colleagues in which the electrons are restricted by nonconstant potential fields to move along the bonds of a periodic network of atoms.[3–5] The network is given the same topological pattern with the same lattice spacings that would represent its crystallographic characterization. If we associate the Schrödinger equation with each bond of the network, then a solution which is satisfied everywhere on the network exists only if the wavefunction and its derivative is everywhere continuous. This statement implies that for the network of atoms depicted in Fig. 2 if $\psi^{ij}(x_{ij})$ is the wavefunction along bond b_{ij}, $0 \leq x_{ij} \leq l_{ij}$, then

$$\psi^{ij_1}_{(0)} = \psi^{ij_2}_{(0)} = \cdots = \psi^{ij_\beta}_{(0)} \tag{2.1a}$$

and

$$\psi'^{ij_1}_{(0)} = \psi'^{ij_2}_{(0)} = \cdots = \psi'^{ij_\beta}_{(0)} \tag{2.1b}$$

where $j_1,\ldots,j_\beta$ is the set of β nodes to which node i is attached.

It should be noted that (2.1a) only holds true for atoms with dimensions of measure zero, i.e., atoms represented as points. Then (2.1b) can be replaced by its graph-theory counterpart, the conservation of flow.[10] Physically, this condition can be thought of as the conservation of probability or momentum at node points and in the stationary state can formally be expressed for arbitrary atomic potential $V(x)$ as[3,11,12]

$$\sum_{(\mathbf{j}_\rho)} \psi'(\mathbf{j}_\rho)\,|_{\mathbf{j}=\text{node}} = 0 \tag{2.2}$$

where the sum is taken over all nearest neighbor node points $\mathbf{j}_\rho$ to $\mathbf{j}$. The energy band structure is then obtained by invoking into (2.2) the appropriate periodic boundary conditions which correspond to the translational symmetry of the lattice. In such manner it can easily be verified that the band structure for a d-dimensional, $1 \leq d \leq 3$, s.c.c. lattice having bond length l can be found from the form factor equation (the equation relating the wavefunction at one node point with the wavefunction at connected node points)

$$2dF(E)\psi(\mathbf{j}) = \sum_{(\mathbf{j}_\rho)} \psi(\mathbf{j}_\rho) \tag{2.3}$$

where the functional form for $F(E)$ is potential dependent, E being the dimensionless energy variable, $E = k^2/2$.[3]

For convenience we choose as our atomic potential

$$V(x) = -V_0 \operatorname{sech}^2 \gamma x \tag{2.4}$$

since this has the property that as the depth of the potential becomes large one obtains a tight binding situation and as the potential vanishes and the number of node points increases, one obtains the Sommerfeld free electron model. γ may be thought of as a force constant and when expressing V_0 as

$$V_0 = r(r+1)(\hbar^2\gamma^2/2m) \tag{2.5}$$

one finds that the wavefunctions, for positive integral values of r, are elementary functions. For example, when $r = 1$

$$\psi(x) = a\left(\cos(kx+\delta) - \frac{\gamma}{k}\sin(kx+\delta)\tanh\gamma x\right) \tag{2.6}$$

a, δ being the constants of integration of the Schrödinger equation. $F(E)$ is then

$$(c^2 - s^2 + cu - sv)/(1 + cu + sv) \tag{2.7}$$

where

$$\tan\theta(x) = \frac{\gamma}{k}\tanh\theta x, \quad \text{with } -\pi < \theta(x) < \pi \tag{2.8a}$$

$$\theta \equiv \theta(l/2) \tag{2.8b}$$

$$c = \cos(kl/2 + \theta) \tag{2.8c}$$

$$s = \sin(kl/2 + \theta) \tag{2.8d}$$

$$u = \left(\frac{\gamma}{k}\right)^2 \cos\theta \operatorname{sech}^2\frac{\gamma l}{2}\cos\frac{kl}{2} \tag{2.8e}$$

and

$$v = \left(\frac{\gamma}{k}\right)^2 \cos\theta \operatorname{sech}^2\frac{\gamma l}{2}\sin\frac{kl}{2}. \tag{2.8f}$$

For k real, Eq. (2.3) will determine the allowable con-

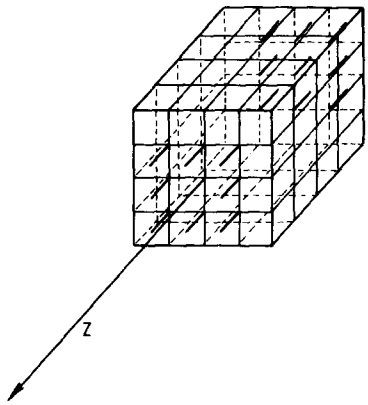


FIG. 3. Surface defect caused by making a cut parallel to the $x-y$ plane along the bond connecting atoms at positions $j_3=1$ and $j_3=N$.

duction band states, while for k imaginary it will determine the bound state band structure.

3. THE LATTICE GREEN'S FUNCTION AND SURFACE DEFECTS

Upon expanding the s. c. c. form factor equation (2. 3) we obtain for

$$\begin{aligned} &\text{1D}\quad 2F(E)\psi(j)=\psi(j\pm1),\\ &\text{2D}\quad 4F(E)\psi(j_1,j_2)=\psi(j_1\pm1,j_2)+\psi(j_1,j_2\pm1),\\ &\text{3D}\quad 6F(E)\psi(j_1,j_2,j_3)=\psi(j_1\pm1,j_2,j_3)+\psi(j_1,j_2\pm1,j_3)\\ &\qquad\qquad +(j_1,j_2,j_3\pm1). \end{aligned} \tag{3.1}$$

The (±) refers to terms $(j+1)$ and $(j-1)$.

By invoking cyclic boundary conditions $\psi(\mathbf{j}+\mathbf{N})=\psi(\mathbf{j})$, where $\mathbf{N}$ is the triple (N_1,N_2,N_3) denoting the number of atoms along the unit vector directions $(\underset{\sim}{i},\underset{\sim}{j},\underset{\sim}{k})$, the complete set of equations for every point $\mathbf{j}$ in our lattice may be written in matric form

$$M\Psi=0 \tag{3.2}$$

where M is the connectivity matrix whose elements show how $\psi(\mathbf{j})$ is associated to $\psi(\mathbf{j}_\alpha)$ via Eqs. (2. 3) and (3. 1), and Ψ is a column vector containing all the $\psi(\mathbf{j})$ as elements.

For example, the 1D analog of (3. 2) has

$$M=\begin{pmatrix} 2F & -1 & 0 & \cdots & -1\\ -1 & 2F & -1 & \cdots & \\ 0 & \cdots & \cdots & & \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ -1 & \cdots & & -1 & 2F \end{pmatrix} \tag{3.3a}$$

and

$$\Psi=\begin{pmatrix}\psi(1)\\ \cdot\\ \cdot\\ \cdot\\ \psi(N)\end{pmatrix}. \tag{3.3b}$$

The solution of this set of coupled equations is gotten by setting $\det|M|=0$. In general, M is a circulant for s. c. c. lattices of any dimension with periodic boundary conditions; thus, the elements of Ψ must be of the form

$$\begin{aligned} &\psi_{\{s\}}(\mathbf{j})=A\exp\left[2\pi i\sum_{\{s\}}\mathbf{s}\cdot\mathbf{j}\right],\\ &\{s\}=s_1,s_2,s_3,\quad s_i=\frac{1}{N_i},\frac{2}{N_i}\cdots 1, \end{aligned} \tag{3.4}$$

with A being a normalization constant. Substitution of (3. 4) into Eqs. (3. 1), in the limit as $N_i\to\infty$, yields the d-dimensional s. c. c. dispersion relation

$$2dF(E)=2\sum_{i=1}^{d}\cos\phi_i, \tag{3.5}$$

$\boldsymbol{\phi}$ being the Brillouin zone wave vector with distribution in $-\pi\le\boldsymbol{\phi}\le\pi$.

One may destroy the periodicity of our cubic lattice by changing atomic masses or potentials, by altering bond lengths, or by creating a "surface." Rewriting M as M_0-M_1 where M_0 is the connectivity matrix of the perfect crystal, and M_1 is the matrix containing the perturbative elements (defects) then (3. 2) becomes

$$\begin{aligned} M\Psi&=(M_0-M_1)\Psi\\ &=0 \end{aligned} \tag{3.6a}$$

and

$$\begin{aligned} \det M&=\det|M_0-M_1|\\ &=\det M_0\det|I-M_0^{-1}M_1| \quad \text{if } \det M_0\neq0. \end{aligned} \tag{3.6b}$$

Premultiplying (3. 6a) by M_0^{-1} leads to the solution of the defect problem

$$\Psi=M_0^{-1}M_1\Psi. \tag{3.7}$$

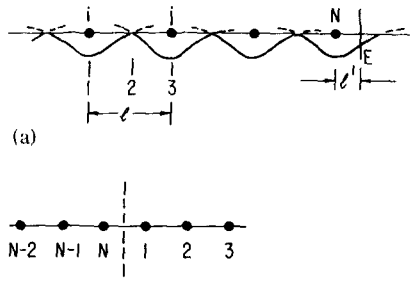
M_0^{-1} will be identified as the lattice Green's function in Appendix A. If the number of defect locations in the lattice is small as compared with the total number of node points, the s. c. c. lattice Green's function may be used to determine the effect of these perturbations on the electronic energy band structure and density of states via (3. 6b).

The s. c. c. Green's function of dimension d is

$$g(\mathbf{j})=\frac{1}{2\mathbf{N}}\sum_{\{s\}=1/\mathbf{N}}^{1}\frac{\exp[2\pi i(\sum_{\{s\}}\mathbf{s}\cdot\mathbf{j})]}{dF(E)-(\sum_{\{s\}}\cos2\pi s)} \tag{3.8}$$

where the $g(\mathbf{j})\equiv M_0^{-1}(\mathbf{j})$ are the elements of M_0^{-1}. This function has been used extensively by Montroll and his co-workers[5] to study a variety of defect problems, among those the interstitial and surface defect. A surface defect is caused by making a cut parallel to the xy plane along the bond connecting atoms at positions $j_3=1$ and $j_3=N_3$, as depicted in Fig. 3. Thus, the periodicity of the structure has only been destroyed along the j_3 direction. Due to this special symmetry the method used for solving the scattering problem from a periodic surface, as discussed in Sec. 5, will be very similar to the scattering from the end atom of a cut circular chain.

Consider a 1D ring comprised of N atoms with a periodic atomic potential of the form $V(x)=-V_0\,\mathrm{sech}^2\gamma x$, V_0 and γ being parameters determining the shape of the potential well, and create a defect by making a cut be-



(a)

(b)

FIG. 4. (a) Schematic drawing of potential $V(x) = -2\gamma^2 \operatorname{sech}^2\gamma x$ applied at each node point. Wave functions and their derivatives much be continuous at connecting points such as 2. (b) Two free ends are created by cutting the bond between N and 1.

tween atom 1 and atom N, as illustrated in Fig. 4. M, thus has the form

$$M = \begin{pmatrix} 2F-\eta-1 & & & 0 \\ -1 & 2F-1 & & 0 \\ \vdots & & & \\ \cdots & & 2F & \cdots \\ \vdots & & & \\ \cdots & & & \cdots \\ \vdots & & & \\ 0 & \cdots & -1 & 2F-\eta \end{pmatrix} \tag{3.9}$$

$$= \begin{pmatrix} 2F & -1 & 0 & \cdots & -1 \\ -1 & 2F & -1 & \cdots & \\ \vdots & & & & \\ \cdots & & & & \cdots \\ \vdots & & & & \\ -1 & \cdots & & -1 & 2F \end{pmatrix} - \begin{pmatrix} \eta & \cdots & -1 \\ \vdots & & \vdots \\ -1 & \cdots & \eta \end{pmatrix}$$

$$= M_0 - M_1.$$

η is a memory function depending upon $V(x)$ and the details of the cut. $F(E)$ is the F of Eq. (2.3). Both functions are derived in Appendix B. From (3.6b)

$$\det(I - M_0^{-1}M_1)$$

$$= \det \begin{pmatrix} 1-\eta g(0)+g(1) & 0 & \cdots & g(0)+\eta g(1) \\ g(2)-\eta g(1) & 1 & & g(1)-\eta g(2) \\ 0 & & 1 & \\ \vdots & & & \vdots \\ -\eta g(1)+g(0) & \cdots & \cdots & 1+g(1)-\eta g(0) \end{pmatrix} = 0.$$

$$= \det \begin{pmatrix} 1-\eta g(0)+g(1) & g(0)-\eta g(1) \\ -\eta g(1)+g(0) & 1+g(1)-\eta g(0) \end{pmatrix} = 0. \tag{3.10}$$

This leads to the coupled equations

$$\begin{aligned} (1+\eta)[g(0)-g(1)] &= 1, \\ (\eta-1)[g(0)+g(1)] &= 1, \end{aligned} \tag{3.11}$$

from which the localized surface state energy structure is determined.

Substitution of the result

$$M_1\Psi = \begin{pmatrix} \eta & -1 \\ \vdots & \vdots \\ -1 & \eta \end{pmatrix} \begin{pmatrix} \psi(1) \\ \vdots \\ \psi(N) \end{pmatrix} = \begin{pmatrix} \eta\psi(1)-\psi(N) \\ 0 \\ \vdots \\ 0 \\ \eta\psi(N)-\psi(1) \end{pmatrix} \tag{3.12}$$

into (3.6a) leads to the defect wavefunction

$$\psi(j) = g(j-1)[\eta\psi(1)-\psi(N)] + g(N-j)[\eta\psi(N)-\psi(1)] \tag{3.13}$$

subject to the condition that $g(j+N) = g(j) = g(-j)$, and where the subscript s has been dropped from the nodal wavefunction.

That the 3D solution has exactly the same form as (3.12) can be demonstrated in the following manner. Creation of the surface previously described results in an equation satisfied by the components of (3.6a) of the form

$$\begin{aligned} &6F(E)\psi(j_1,j_2,j_3) - \psi(j_1\pm1,j_2,j_3) - \psi(j_1,j_2\pm1,j_3) \\ &\quad - \psi(j_1,j_2,j_3\pm1) = G(j_1,j_2,j_3) \end{aligned} \tag{3.14a}$$

where

$$\begin{aligned} Gj_1,j_2,j_3) &= \delta_{j_3,N_3}[\eta\psi(j_1,j_2,N_3) - \psi(j_1,j_2,1)] \\ &\quad + \delta_{j_3,1}[\eta\psi(j_1,j_2,1) - \psi(j_1,j_2,N_3)]. \end{aligned} \tag{3.14b}$$

Due to the surface periodicity $\psi(j_1,j_2,j_3)$ may be rewritten as

$$\psi(j_1,j_2,j_3) = A\exp\left[2\pi i\left(\frac{j_1n_1}{N_1} + \frac{j_2n_2}{N_2}\right)\right]\phi(j_3). \tag{3.15}$$

By inserting this into the Eqs. (3.14) we obtain

$$\begin{aligned} &2\left(3F - \frac{\cos2\pi n_1}{N_1} - \frac{\cos2\pi n_2}{N_2}\right)\phi(j_3) - \phi(j_3\pm1) \\ &\quad = \delta_{j_3,N_3}[\eta\phi(N_3) - \phi(1)] + \delta_{j_3,1}[\eta\phi(1) - \phi(N_3)] \end{aligned} \tag{3.16}$$

whose solution due to (3.7) is

$$\begin{aligned} \phi(j_3) &= \sum_{j_3'} g(j_3-j_3')G(j_3') \\ &= g(j_3-N_3)[\eta\phi(N_3) - \phi(1)] + g(j_3-1)[\eta\phi(1) - \phi(N_3)] \end{aligned} \tag{3.17}$$

where

$$g(j_3) = \frac{1}{2N_3}\sum_{s_3=1}^{N_3} \frac{\exp(2\pi i s_3 j_3/N_3)}{F_{n_1,n_2} - \cos2\pi s_3/N_3} \tag{3.18a}$$

and

$$F_{n_1,n_2} = 3F - \frac{\cos2\pi n_1}{N_1} - \frac{\cos2\pi n_2}{N_2}. \tag{3.18b}$$

To conclude this section, a few comments about periodic lattices, in general, should be made. As long as a form factor equation resembling (2.3) can be derived, the matric equation $M\Psi = 0$ can be solved, thereby determining the elements of Ψ. For a periodic lattice Ψ will always contain the translational symmetry of the reciprocal lattice of the space group of the structure. The Green's function can then be found by inverting M_0. Green's functions for body and face centered cubic lattices with coordination numbers of eight and

twelve, respectively, are

3D body centered

$$g(j_1,j_2,j_3)=\frac{1}{8N_1N_2N_3}\sum_{s_1=1}^{N_1}\sum_{s_2=1}^{N_2}\sum_{s_3=1}^{N_3}\frac{\exp[2\pi i(\sum s_l j_l)/N_l)}{F(E)-\Pi_1^3(\cos 2\pi s_i/N_i)}$$

3D Face centered (3.19)

$$g(j_1,j_2,j_3)=\frac{1}{4N_1N_2N_3}\sum_{\{s\}=1}^{N_1,N_2,N_3}\frac{\exp[2\pi i(\sum s_l j_l)/N_l]}{3F(E)-[(\cos 2\pi s_1/N_1)(\cos 2\pi s_2/N_2)+(\cos 2\pi s_1/N_1)(\cos 2\pi s_3/N_3)+(\cos 2\pi s_2/N_2)(\cos 2\pi s_3/N_3)]}.$$

4. SCATTERING OF BLOCH ELECTRONS FROM DEFECTS

In Sec. 3 it was shown that the basic equation from which all the electronic properties of a crystal come is

$$M\Psi=0. \tag{4.1}$$

Scattering of Bloch electrons from a crystal surface or from some other internal defect in the crystal may be studied by setting $\psi=\psi_0+\omega$, where ψ_0 is an incoming wave and a solution to the perfect crystal matrix form factor equation, i.e., $M_0\Psi_0=0$, and ω is a scattered wave. Then (4.1) with the auxiliary perfect crystal condition becomes

$$M\Psi=(M_0-M_1)(\Psi_0+\Omega)=0 \tag{4.2}$$

or

$$M_1\Psi_0=-M_1\Omega+M_0\Omega.$$

Premultiplying the second equation of (4.2) by M_0^{-1} yields

$$M_0^{-1}M_1\Psi_0=(I-M_0^{-1}M_1)\Omega \tag{4.3}$$

or

$$\Omega=M_0^{-1}M_1\Psi_0+M_0^{-1}M_1\Omega$$

and by iteration

$$=M_0^{-1}M_1\Psi_0+(M_0^{-1}M_1)^2\Psi_0+(M_0^{-1}M_1)^3\Psi_0+\cdots.$$

This series is analogous to the Born expansion to Ω. A formally exact solution may be written as

$$\Omega=M_0^{-1}T\Psi_0 \tag{4.4a}$$

where

$$T=M_1+M_1M_0^{-1}T. \tag{4.4b}$$

The T matrix in (4.4b), or transition matrix, is to be identified with the well-known T matrix of collision theory. Since this T-matrix equation contains the defect matrix M_1, the scattering from the defect to any other lattice point is a multiple scattering phenomenon.

We now show that the scattered wave ω asymptotically resembles an outgoing wave. This discussion will be limited to the cubic lattices whose Green's functions are given by (3.8) and (3.19), although the general flavor of this argument is also applicable to lattices of other symmetry. The general cubic lattice Green's function can be expressed as

$$g(j_1,j_2,j_3)=\frac{1}{c_1N_1N_2N_3}\times\sum_{s_1,s_2,s_3}\frac{\exp[2\pi i(s_1j_1/N_1+s_2j_2/N_2+s_3j_3/N_3)]}{c_2F(E)-\lambda(2\pi s_1/N_1,2\pi s_2/N_2,2\pi s_3/N_3)} \tag{4.5}$$

where for

s.c.c.

$$c_1=2,\quad c_2=3,$$

$$\lambda(\phi)=\cos\left(\frac{2\pi s_1}{N_1}\right)+\cos\left(\frac{2\pi s_2}{N_2}\right)+\cos\left(\frac{2\pi s_3}{N_3}\right)$$

b.c.c.

$$c_1=8,\quad c_2=1$$

$$\lambda(\phi)=\cos\frac{2\pi s_1}{N_1}\cos\frac{2\pi s_2}{N_2}\cos\frac{2\pi s_3}{N_3}$$

f.c.c.

$$c_1=4,\quad c_2=3,$$

$$\lambda(\phi)=\cos\frac{2\pi s_1}{N_1}\cos\frac{2\pi s_2}{N_2}+\cos\frac{2\pi s_1}{N_1}\cos\frac{2\pi s_3}{N_3}+\cos\frac{2\pi s_2}{N_2}\cos\frac{2\pi s_3}{N_3}.$$

Defining $\mathbf{x}(j)=j_1\mathbf{a}_1+j_2\mathbf{a}_2+j_3\mathbf{a}_3$ as the position vector of the jth atom in the lattice, $\mathbf{a}_i$ being the primitive translation vectors of the lattice, and

$$\mathbf{q}=2\pi\left(\frac{s_1\mathbf{b}_1}{N_1}+\frac{s_2\mathbf{b}_2}{N_2}+\frac{s_3\mathbf{b}_3}{N_3}\right)$$

where the $\mathbf{b}_i$ are the primitive translation vectors of the reciprocal lattice, such that $\mathbf{a}_i\cdot\mathbf{b}_j=\delta_{ij}$ and $s_i=1,2,\ldots,N_i$, then

$$\exp[i\mathbf{q}\cdot\mathbf{x}(j)]=\exp(2\pi i\mathbf{s}\cdot\mathbf{j}/\mathbf{N}). \tag{4.6}$$

Thus, (4.5) may be rewritten as

$$g(j_1,j_2,j_3)=\lim_{\delta\to0}\frac{1}{c_1N_1N_2N_3}\times\sum_{s_1,s_2,s_3}\frac{\exp(i\mathbf{q}\cdot\mathbf{x}(j)]}{c_2F(E)-\lambda(2\pi\mathbf{s}/\mathbf{N})+i\delta}. \tag{4.7}$$

Using the integral representation of

$$\frac{1}{x+i\delta}=\frac{1}{i}\int_0^\infty dt\exp(ixt-\delta t) \tag{4.8}$$

and noting that for an asymptotic $g(j_1,j_2,j_3)$ one can change the sum over s in (4.6) to an integral over ϕ, where the allowed values of ϕ lie in the first Brillouin zone of the crystal, i.e., $-\pi\le\phi\le\pi$, then (4.7)

becomes

$$g(j_1,j_2,j_3)=\frac{1}{ic_1(2\pi)^3}\int_{Bz}d^3\phi\int_0^\infty dt\exp[iG(E,\phi)] \tag{4.9a}$$

where

$$G(E,\phi)=\mathbf{q}\cdot\mathbf{x}(j)+[c_2F(E)-\lambda(\phi)]t \tag{4.9b}$$

and

$$\phi_i=2\pi s_i/N_i. \tag{4.9c}$$

For large $x(j)$ the exponential under the integral will oscillate quite rapidly. By using the method of stationary phase the greatest contribution from the integral can be determined.

Setting

$$\frac{\partial G}{\partial t}=c_2F(E)-\lambda(\phi)=0 \tag{4.10}$$

and

$$\frac{\partial G}{\partial\phi}=\mathbf{j}-\frac{\partial\lambda(\phi)t}{\partial\phi}=\mathbf{j}-\frac{\partial\lambda}{\partial E}\frac{\partial E}{\partial\phi}t=0, \tag{4.10}$$

one finds that the largest contribution to the integral is along the surface of constant energy and around the neighborhood of a point where the electron group velocity is parallel and in the same direction to the position vector. Thus, the scattering process occurs in the energy bands (Bloch electrons) of the material. We label the values of t and ϕ for which this occurs t_m and ϕ_m.

Expanding $G(E,\phi)$ about the point (t_m,ϕ_m) gives

$$\begin{aligned}G(E,\phi)=\phi_m&-(t-t_m)\sum_{i=1}^{3}(\phi_i-\phi_{im})\frac{\partial\lambda(\phi_i)}{\partial\phi_i}\Big|_{\phi_i=\phi_{im}}\\&-\tfrac{1}{2}t_m\sum_{i,j}(\phi_i-\phi_{im})(\phi_j-\phi_{jm})\\&\times\frac{\partial^2\lambda(\phi)}{\partial\phi_i\,\partial\phi_j}\Big|_{\substack{\phi_i=\phi_{im}\\ \phi_j=\phi_{jm}}}+\cdots.\end{aligned} \tag{4.11}$$

All first order terms have dropped out due to (4.10).

Introducing the relative variables $\tau=t-t_m$, $\chi=\phi_m-\phi_m$ and extending the range of integration for all variables in (4.9a) over all space yields

$$\begin{aligned}g(j_1,j_2,j_3)\sim\frac{\exp(i\phi_m)}{ic_1(2\pi)^3}&\int\int_{-\infty}^{+\infty}\int\int d\chi\,d\tau\\&\times\exp[-i\tau(\nabla_\phi\lambda(\phi_m)\cdot\chi)]\exp(-\tfrac{1}{2}it_m[\cdots]).\end{aligned} \tag{4.12}$$

The brackets in the exponent contain the third term of $G(E,\phi)$. One chooses a coordinate system for χ so that one coordinate (χ_z) is perpendicular to the surface of constant energy at $\phi=\phi_m$, i.e., parallel to $\partial E/\partial\phi|_{\phi=\phi_m}$ and the other two so that the cross terms $\chi_x\chi_y$ in the quadratic form in the exponential of (4.12) vanishes.

Performing the integration over τ gives

$$\begin{aligned}g(j_1,j_2,j_3)\sim\frac{\exp(i\phi_m)}{ic_1(2\pi)^2}\cdot\frac{1}{|\nabla_\phi\lambda(\phi_m)|}&\int_{-\infty}^{+\infty}d\chi_x\int_{-\infty}^{+\infty}d\chi_y\\&\times\exp(-it_m[\chi\cdot\nabla_\phi\nabla_\phi\cdot\chi]/2)\end{aligned} \tag{4.13}$$

where the vectors χ and ∇_ϕ contained in the brackets of the exponential are now two-dimensional vectors.

Using the identity

$$\int_{-\infty}^{+\infty}\exp(-i\lambda x^2)\,dx=\frac{\pi^{1/2}}{|\lambda|^{1/2}}\exp[-(i\pi/4)\operatorname{sgn}\lambda] \tag{4.14}$$

leads us to the final result

$$\begin{aligned}g(j_1,j_2,j_3)&\sim\frac{\exp(i\phi_m)}{2\pi|\mathbf{X}(j)|}\\&\times\frac{\exp(-i\pi\{(\operatorname{sgn}/4)[\partial^2\lambda(\phi_{mx})/\partial\chi_x^2]+(\operatorname{sgn}/4)[\partial^2\lambda(\phi_{my})/\partial\chi_y^2]+\tfrac{1}{2}\})}{|\partial^2\lambda(\phi_{xm})/\partial\chi_x^2|^{1/2}|\partial^2\lambda(\phi_{ym})/\partial\chi_y^2|^{1/2}}\end{aligned} \tag{4.15}$$

At times the energy lying on the interior of the surface of constant energy is slightly higher than the energy lying on the exterior. In such a case the vector $\nabla_\phi\lambda(\phi)$ points inward, and the inner product of $\mathbf{q}_m\cdot\mathbf{X}(j)<0$ implying that (4.13) will correspond to an incoming wave rather than an outgoing wave. This whole analysis may be saved if instead of (4.7) our Green's function is chosen to be

$$g(j_1,j_2,j_3)=\lim_{\delta\to0}\frac{1}{c_1N_1N_2N_3}\sum_{s_1,s_2,s_3}\frac{\exp[i\mathbf{q}\cdot\mathbf{X}(j)]}{c_2F(E)-\lambda(2\pi\mathbf{s}/\mathbf{N})-i\delta}. \tag{4.16}$$

Since

$$\frac{1}{x-i\delta}=-\frac{1}{i}\int_0^\infty dt\exp(-ixt-\delta t), \tag{4.17}$$

by proceeding as in the previous discussion we then will see that our final Green's function will asymptotically resemble an outgoing wave.

5. 1D AND 3D SCATTERING FROM SURFACE DEFECT

As examples of the preceding discussion we now calculate the T matrix and scattered wave ω for a cut 1D circular ring and for a cut 3D s.c.c. periodic lattice as derived from Eqs. (3.7) and (4.4b).

1 *D chain*: Using (3.9),

$$T=\begin{pmatrix}\eta&\cdots&-1\\ \vdots&&\vdots\\ -1&\cdots&\eta\end{pmatrix}+\begin{pmatrix}\eta&\cdots&-1\\ \vdots&&\vdots\\ -1&\cdots&\eta\end{pmatrix}\begin{pmatrix}g(0)&g(1)&\cdots&g(N-1)\\ &&&\vdots\\ g(1)&\cdots&\cdots&g(0)\end{pmatrix}T. \tag{5.1}$$

Noting that the value of the Green's function only depends on the difference in coordinate positions,[5] we may write $g(k-l)\equiv g_{kl}$. Thus, the mrth element of T is

$$\begin{aligned}T_{mr}&=(M_1)_{mr}+\sum_{kl}(M_1)_{mk}g_{kl}T_{lr}\\&=(M_1)_{mr}+\sum_{kl}(\eta[\delta_{m1}\delta_{k1}+\delta_{mn}\delta_{kn}]\\&\quad-[\delta_{mn}\delta_{k1}+\delta_{m1}\delta_{kn}])g_{kl}T_{lr}\\&=(M_1)_{mr}+\eta\sum_l\delta_{m1}g_{1l}T_{lr}+\delta_{mn}g_{nl}T_{lr}\\&\quad-\sum_l\delta_{mn}g_{1l}T_{lr}+\delta_{m1}g_{nl}T_{lr}\end{aligned} \tag{5.2a}$$

where

$$(M_1)_{mr} = \eta[\delta_{m1}\delta_{r1} + \delta_{mn}\delta_{rn}] - [\delta_{mn}\delta_{r1} + \delta_{m1}\delta_{rn}]. \quad (5.2b)$$

Rewriting (5.2a) as

$$T_{mr}[1 + \delta_{mn}g_{1m} + \delta_{m1}g_{nm} - \eta(\delta_{m1}g_{1m} + \delta_{mn}g_{nm})] = (M_1)_{mr} + \eta \sum_{l\neq m} (\delta_{m1}g_{1l}T_{lr} + \delta_{mn}g_{nl}T_{lr}) - \sum_{l\neq m} (\delta_{mn}g_{1l}T_{lr} + \delta_{m1}g_{nl}T_{lr}), \quad (5.2c)$$

it may easily be verified that if $i = 2, \ldots, n-1$ and $j = 1, \ldots, n$, then

$$(M_1)_{ij} = T_{ij} = 0. \quad (5.2d)$$

When $i = 1, \ldots, n$ and $j = 2, \ldots, n-1$,

$$T_{ij} = \frac{[\eta g(1) - g(0)]^2}{[1 - \eta g(0) + g(1)]^2} T_{ij}. \quad (5.2e)$$

This equation is satisfied if $T_{ij} = 0$ or if $[\eta g(1) - g(0)]^2 = [1 - \eta g(0) + g(1)]^2$. That the former relation is the correct one may be demonstrated by noting that

$$T_{11} = T_{nn} = \frac{\eta + g(0)[1 - \eta^2]}{[1 - \eta g(0) + g(1)]^2 - [\eta g(1) - g(0)]^2} \quad (5.2f)$$

and

$$T_{1n} = T_{n1} = \frac{-1 + g(1)[\eta^2 - 1]}{[1 - \eta g(0) + g(1)]^2 - [\eta g(1) - g(0)]^2}. \quad (5.2g)$$

The total wave solution is thus,

$$\psi_s(j) = \exp(2\pi isj/N) + \sum_{\beta\gamma} g_{\alpha\beta}T_{\beta\gamma}\psi_{0_s}(\gamma) = \psi_{0_s}(j) + g(j-1)[T(0)\psi_{0_s}(1) + T(1)\psi_{0_s}(N)] + g(N-j)[T(1)\psi_{0_s}(1) + T(0)\psi_{0_s}(N)] \quad (5.3a)$$

where

$$\psi_{0_s}(j) = \exp(2\pi isj/N). \quad (5.3b)$$

To show the outgoing wave character of the scattered solution, consider the 1D analog of 3.8

$$g(j) = \frac{1}{2N} \sum_{s=1}^{N} \frac{\exp(2\pi isj/N)}{F - \cos 2\pi s/N}. \quad (5.4)$$

As $j \to \infty$

$$g(j) = \frac{1}{4\pi} \int_0^{2\pi} \frac{\exp(ij\phi)\, d\phi}{F - \cos\phi} \quad \text{where } \phi = 2\pi s/N. \quad (5.5)$$

Setting $z = \exp(i\phi)$, (5.5) becomes a contour integral over a unit circle and

$$g(j) = \frac{1}{2\pi i} \oint \frac{z^j\, dz}{z(2F - (z + z^{-1}))} = \frac{-1}{2\pi i} \oint \frac{z^j\, dz}{[z - (F + \sqrt{F^2 - 1})][z - (F - \sqrt{F^2 - 1})]}. \quad (5.6)$$

The value of this integral is dependent upon the range of $|F|$. For,

(1) $|F| > 1$, $g(j) = \dfrac{(|F| - \sqrt{F^2 - 1})^j}{2\sqrt{F^2 - 1}}$; (5.7a)

(2) $|F| < 1$, $g(j) = \dfrac{-(|F| + i\sqrt{1 - F^2})^j}{2i\sqrt{1 - F^2}}$; (5.7b)

(3) $|F| = 1$, $g(j) = -j$. (5.7c)

In case (3) $g(j)$ diverges as $j \to \pm\infty$. This shows the delocalizability of the defect wavefunction at the energy band edge, a result previously obtained by Montroll.[5]

If in (5.7a) one substitutes $\cosh\phi_0$ for $|F|$, one finds

(1) $g(j) = \exp(-j\phi_0)/2\sinh\phi_0$. (5.8a)

Thus, out of the energy band the scattered wave will die out exponentially with the distance from the defect. This, of course, is not a solution to our scattering problem since the incoming wave ψ_0 is not a solution to $M_0\Psi_0 = 0$. Setting $|F| = \cos\phi_0$ in (5.7b) gives

(2) $g(j) = i\exp(ij\phi_0/2\sin\phi_0)$. (5.8b)

Since ϕ_0 varies between $-\pi < \phi_0 < \pi$, our scattered wave will resemble an outgoing or an incoming solution depending on whether j is positive or negative.

The functional form for the scattered Bloch wave will then be given by inserting (5.8b) into (5.3). Due to the surface periodicity a similar result will pertain for the 3D s.c.c. lattice with the replacement of F in (5.4) by F_{n_1,n_2} and with a corresponding reassignment for the nonzero T-matrix elements.

6. SURFACE SCATTERING

The lattice Green's function as defined in Sec. 3 has been utilized to examine the effect of a finite number of defects as perturbative agents on the electronic band structure of a perfectly periodic lattice. Section 4 exploited this construction in order to characterize how a wave propagates from a defect to other points in a solid via a T matrix. It should be clear from Eqs. (4.4) that the lattice Green's function embraces all the multiple-scattering information the T matrix does and therefore by expressing the defect wavefunction by (3.7)

$$\Psi = M_0^{-1}M_1\Psi \quad (3.7)$$

we obtain a solution to the defect problem which implicitly contains the combined multiple scattering properties of the periodic lattice together with its effective defect(s).

Imagine an incident particle, either an electron or photon, striking the surface of our crystal system and then being scattered back into free space. We restrict ourselves to just this first layer scattering because the surface defect "communicates" with the bulk of the crystal through the lattice Green's function. One could make the model more realistic by including the second and/or the third layer as defects, but this would only change the functional form of $G(j_1, j_2, j_3)$ in (3.14a) leaving Eq. (3.7) structurally invariant.

Utilizing Eqs. (3.15), (3.17), and (3.18) the surface wavefunction in the limit as $N_3 \to \infty$ for fixed j_3 is

$$\psi(j_1, j_2, j_3) = \exp\left[2\pi i\left(\frac{n_1 j_1}{N_1} + \frac{n_2 j_2}{N_2}\right)\right](\eta - 1)g(j_3 - 1)\phi(1). \quad (6.1)$$

Since $g(j_3)$ may have localized or propagating character depending on whether $|F_{n_1,n_2}| > 1$ or $|F_{n_1,n_2}| < 1$, see Eqs. (5.7) respectively, $\psi(j_1, j_2, j_3)$ when normalized over all the crystal atoms has the form

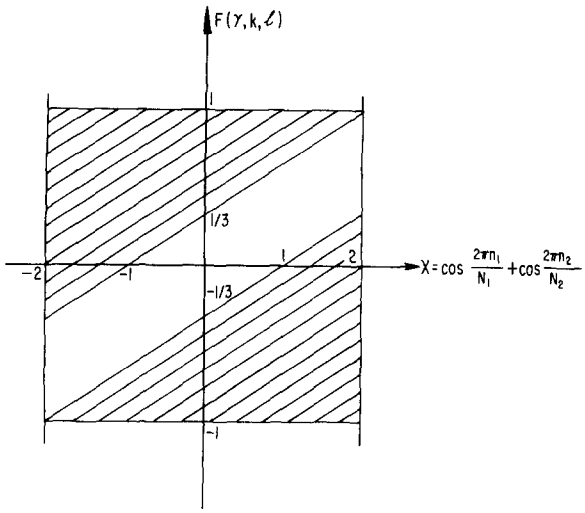


FIG. 5. Plot of the form factor $F(k/\gamma)$ as a function of $X(2\pi n_1/N_1,\ 2\pi n_2/N_2)$ for surface states. The shaded area in the range $|F|<1$ is a region in which the surface energy levels overlap the bulk energy bands. In the shaded area for which $|F|>1$, the surface states lie outside the band of bulk energy bands.

$$\psi(j_1,j_2,j_3) = \begin{cases} (N_1N_2)^{-1/2}\exp[i(j_1\theta_1+j_2\theta_2)][1-\exp(-2Z_{n_1,n_2})]^{1/2} \\ \quad\times \mathrm{sgn}(F_{n_1,n_2})\exp[-(j_3-1)Z_{n_1,n_2}] \qquad (6.2a) \\ i(N_1N_2N_3)^{-1/2}\exp[i(j_1\theta_1+j_2\theta_2)]\,\mathrm{sgn}(\tilde{F}_{n_1,n_2}) \\ \quad\times\exp[i(j_3-1)\tilde{Z}_{n_1,n_2}] \qquad (6.2b) \end{cases}$$

where we have made the substitutions

$$\theta_i = 2\pi n_i/N_i, \qquad (6.3)$$

$$|F_{n_1,n_2}| = \cosh Z_{n_1,n_2}, \qquad (6.4a)$$

$$U_{n_1,n_2} = \mathrm{sgn}(F_{n_1,n_2})\exp(-Z_{n_1,n_2}), \qquad (6.4b)$$

$$|\tilde{F}_{n_1,n_2}| = \cos\tilde{Z}_{n_1,n_2}, \qquad (6.5a)$$

and

$$\tilde{U}_{n_1,n_2} = \mathrm{sgn}(\tilde{F}_{n_1,n_2})\exp(i\tilde{Z}_{n_1,n_2}). \qquad (6.5b)$$

Figure 5 shows a plot of the form factor $F(E)$ as a function of $X(2\pi n_1/N_1, 2\pi n_2/N_2)$. For $|F|<1$, the surface states overlap the bulk energy bands. Since the bulk and surface wavefunctions are continuous at the surface, the net result of this is that when an incident particle scatters from surface states whose energies lie in the bulk bands they are in essence scattering from bulk states as well. For $|F|>1$, the incident particle will scatter solely from Tamm states.

The actual details for the computation of the scattering intensity are now no more difficult than the matching of a number of wavefunction boundary conditions. Since the atomic potentials of our crystal are all 1D potentials, the impinging particles can only interact with the field of an atom scatterer along the direction of a bond. For simplicity, assume that the scattering in a center of mass coordinate system takes place along the dangling bonds of our surface (see Fig. 6). The potential $V(r)$, $0\le r\le l'$, along that bond may be given 3D character by setting $r=z\cos\theta$. $z\cos\theta$ is therefore a projection of r with θ defining the incident scattering angle. Thus, when the scattering angle $\theta=0$ (direct incidence) the dangling bond is normal to the surface, i.e., its direction when the surface was created, and the incident particle feels the full potential field of the surface atom. At any angle $\theta\neq 0$ the interaction is correspondingly lessened. Since the crystal energy structure is of topological origin, changing the orientation of the dangling surface bonds over a polar angle θ, $-\pi/2\le\theta\le\pi/2$, does not affect the crystal energy structure.

A solution to this external scattering problem can be obtained by postulating a surface potential that is cut off at some distance $r=R$ from the scatterer. Since this cutoff potential actually coincides with the end of the dangling bond, we have

$$V(r) = \begin{pmatrix} -V_0\,\mathrm{sech}^2\gamma r & 0\le r\le l' \\ 0 & r>l' \end{pmatrix}. \qquad (6.6)$$

The solution to the Schrödinger equation thus has three regions of validity

$$(1)\quad r=0,\quad \phi(0)=\psi(j_1,j_2,1)=c\cos\delta = \frac{\exp[2\pi i(N_1j_1/N_1+n_2j_2/N_2}{(N_1N_2)^{1/2}} \times(1-U^2_{n_1,n_2})^{1/2}, \qquad (6.7a)$$

$$(2)\quad 0\le r\le l',\quad \phi(r)=c\cos\delta\left(\cos kr-\frac{\gamma}{k}\sin kr\tanh\gamma r - \tan\delta\left(\sin kr+\frac{\gamma}{k}\cos kr\tanh\gamma r\right)\right), \qquad (6.7b)$$

$$(3)\quad r>l',\quad \phi(r)=B\exp(+ik_0r), \qquad (6.7c)$$

where k_0^2 is the incident particle energy, $k_0^2=k^2+2\gamma^2$.

In (6.7a) we have chosen the surface wavefunction corresponding to localized Tamm states since it has the damping structure necessary to mitigate the penetrating power of the incident particle to the first few levels of the crystal. The ensuing calculation could also be performed for surface wavefunctions of the type (6.2b). By the continuity of the wavefunction and its derivative at $r=l'$, and due to (6.7a), it can be shown after some algebra that

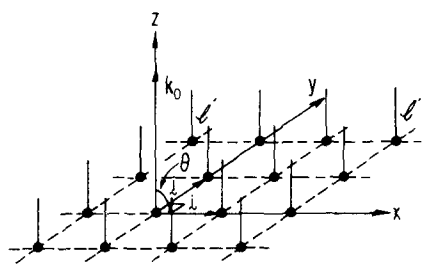


FIG. 6. A rectangular two-dimensional lattice of scatterers: $j_1\underset{\sim}{i}$ and $j_2\underset{\sim}{j}$ are the primitive lattice vectors. The incident wavevector $\mathbf{k}_0$ is shown perpendicular to the lattice plane, with θ defining the incident scattering angle.

$$\tan\delta = \frac{-(\gamma^2/k^2)\operatorname{sech}^2\gamma l'(\sin 2kl' + (\gamma/k)\tanh\gamma l' + (\gamma^2/2k^2)\operatorname{sech}^2\gamma l'\sin 2kl') - i(1+(\gamma^2/k^2)(\operatorname{sech}^2\gamma l' + \tanh^2\gamma l'))}{1+(\gamma^2/k^2)(\tanh^2\gamma l' + 2\cos^2 kl'\operatorname{sech}^2\gamma l'(1+(\gamma^2/2k^2)\operatorname{sech}^2\gamma l') - (\gamma/k)\sin 2kl'\tanh\gamma l'\operatorname{sech}^2\gamma l')}, \quad (6.8a)$$

$$\lim_{l'\to\infty}\tan\delta = -i, \quad (6.8b)$$

and

$$B = \frac{\exp(-ik_0 l')}{(N_1N_2)^{1/2}}\exp\left[2\pi i\left(\frac{m_1 j_1}{N_1}+\frac{m_2 j_2}{N_2}\right)\right](1-U^2_{m_1,m_2})^{1/2}\left(\cos kl' - \frac{\gamma}{k}\sin kl'\tanh\gamma l' - \tan\delta\left(\sin kl' + \frac{\gamma}{k}\cos kl'\tanh\gamma l'\right)\right). \quad (6.8c)$$

If each surface scatterer acts independently of one another then the total scattered wave $\Phi(P)$ at some point P above the surface due to contributions from all the scattering centers on the lattice surface is a superposition of plane waves. Realistically, though, this constructive interference can occur only when the point P is asymptotically distant from the surface. Thus,

$$\Phi(P) = B\sum_{j_1,j_2=1}^{N_1,N_2}\phi_{j_1,j_2}(P) = B\sum_{j_1,j_2=1}^{N_1,N_2}\exp(ik_0|r_{j_1,j_2}|) \quad (6.9)$$

where the sum is over all the coherent scattered waves emitted from scattering centers which are located at coordinates $(j_1, j_2, 1)$

$$j_1 = 1, \ldots, N_1,$$

$$j_2 = 1, \ldots, N_2$$

and where we identify $P \equiv |r_{j_1,j_2}|$.

We sum the series in (6.9) by considering a vector κ pointing in the same direction as P and with magnitude equal to k_0. Then,

$$\begin{aligned}\kappa\cdot\mathbf{r}_{j_1,j_2} &= |\kappa|\,|r_{j_1,j_2}| \\ &= k_0|r_{j_1,j_2}| \\ &= k_0 P.\end{aligned} \quad (6.10)$$

Due to the surface periodicity, any point may be chosen as an origin (r') with the vector $\mathbf{r}_{j_1,j_2}$ joining the point P above the surface with r' on the surface. Defining $\mathbf{u} = \mathbf{r}' - \rho(j_1,j_2)$ as a vector connecting any point on the lattice surface with the origin at r', $\mathbf{r}_{j_1,j_2}$ may be rewritten as

$$\mathbf{r}_{j_1,j_2} = (\mathbf{p} - \mathbf{u} - \rho(j_1,j_2));$$

(6.9) then becomes

$$B\exp[i\kappa\cdot(\mathbf{p}-\mathbf{u})]\sum_{j_1,j_2}\exp[-i\kappa\cdot\rho(j_1,j_2)]. \quad (6.11)$$

Defining

$$\rho(j_1,j_2) = (j_1\underset{\sim}{i}, j_2\underset{\sim}{j}, 0\underset{\sim}{k}) \quad (6.12a)$$

and

$$\kappa = k_0(\underset{\sim}{i},\underset{\sim}{j},\underset{\sim}{k}), \quad (6.12b)$$

the triple $(\underset{\sim}{i},\underset{\sim}{j},\underset{\sim}{k})$ denoting the standard Cartesian unit vectors, (6.11) then sums to

$$B\exp[i\kappa\cdot(\mathbf{p}-\mathbf{u})]\left(\frac{1-\exp(iN_1k_{0x})}{1-\exp(ik_{0x})}\right)\left(\frac{1-\exp(iN_2k_{0y})}{1-\exp(ik_{0y})}\right). \quad (6.13)$$

By setting $N_1 = N_2 = N$ the scattered intensity $I(\theta,\phi)$ is then

$$\begin{aligned}I(\theta,\phi) &= |\Phi(P)|^2 \\ &= |B|^2\frac{\sin^2(Nk_{0x}/2)}{\sin^2(k_{0x}/2)}\frac{\sin^2(Nk_{0y}/2)}{\sin^2(k_{0y}/2)}\end{aligned} \quad (6.14)$$

where

$$k_{0x} = k_0\sin\theta\cos\phi,$$

$$k_{0y} = k_0\sin\theta\sin\phi,$$

and ϕ is the azimuthal angle of spherical coordinates. B in this formulation may be identified with the geometrical structure factor[13] of Bragg diffraction, since it clearly depends upon the connectivity and atomic potentials of the surface. Functionally speaking Eq. (6.14) resembles the intensity found due to Fraunhofer diffraction in optical systems.[14]

For illustrative purposes, suppose $\phi = \pi/4$. Then, I resembles

$$\sin^4 Nx/\sin^4 x. \quad (6.15)$$

This function (Fig. 7) has maxima, each of height N^4, at $x = 0$ and at $x = n\pi$, $n = \pm1, \pm2, \cdots$; $x = 0$ may be obtained by either setting $k_0 = 0$, or $\theta = 0$. Physically, these situations may be interpreted as the forward scattering due to the incident particle–valence electron interaction, since this process is greatest when these conditions are met. Secondary maxima occur between the zeroes of the intensity function, that is, when $x = m\pi/N$, m/N nonintegral, and where $m = \pm1, \pm2, \cdots$. For $\phi \neq \pi/4$ similar analyses may be performed, resulting in the same type of qualitative behavior for I.

As $N \to \infty$

$$I(\theta,\phi) \sim N^4\delta(k_{0x} - 2\pi m_1)\delta(k_{0y} - 2\pi m_2)_{m_1,m_2=0,\pm1,\pm2,\cdots}. \quad (6.16)$$

Since $k_0 = 2\pi/\lambda$, then when the wavelength of the incoming radiation λ is of the same order as the interatomic spacing of the surface; this relation expresses the well-known condition for Bragg scattering in terms of θ and ϕ.

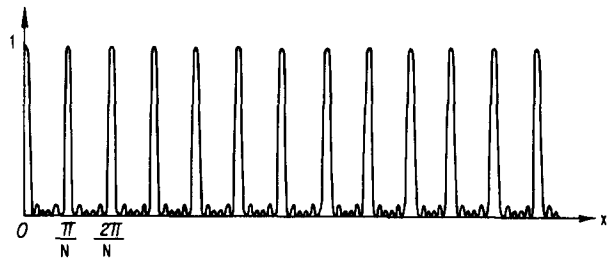
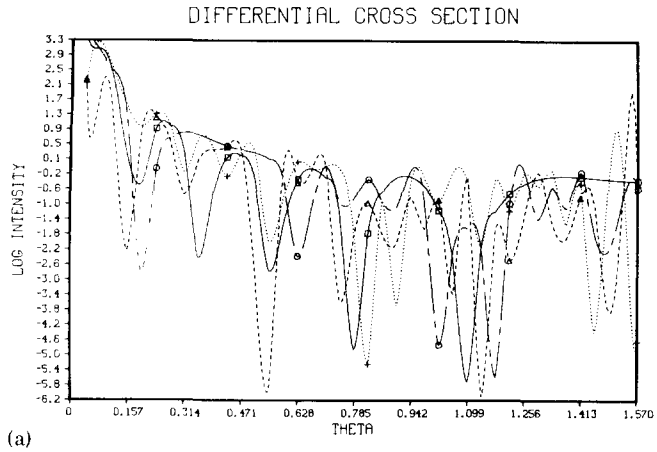
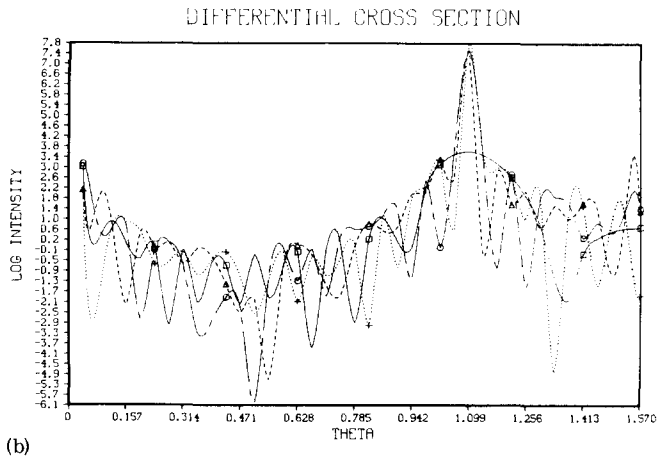


FIG. 7. The normalized intensity function

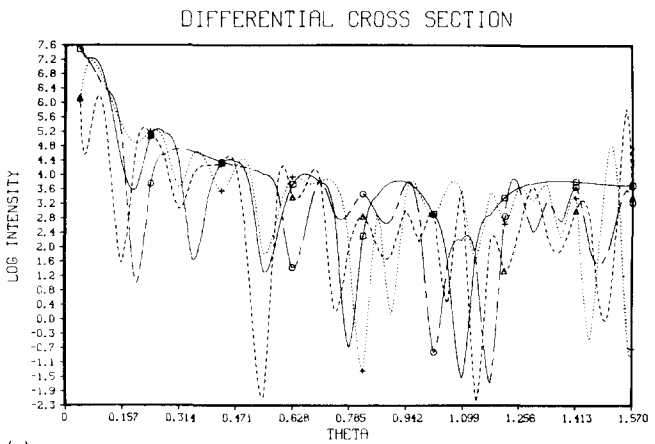
$$I(\theta,\pi/4) = \frac{1}{N^4}\frac{\sin^4 Nx}{\sin^4 x}.$$



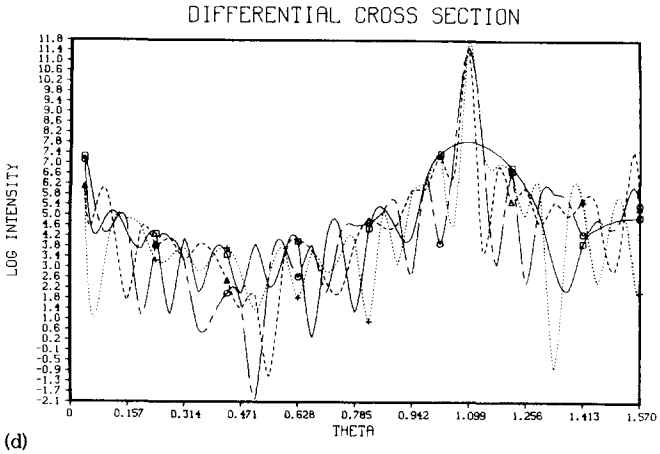
(a)



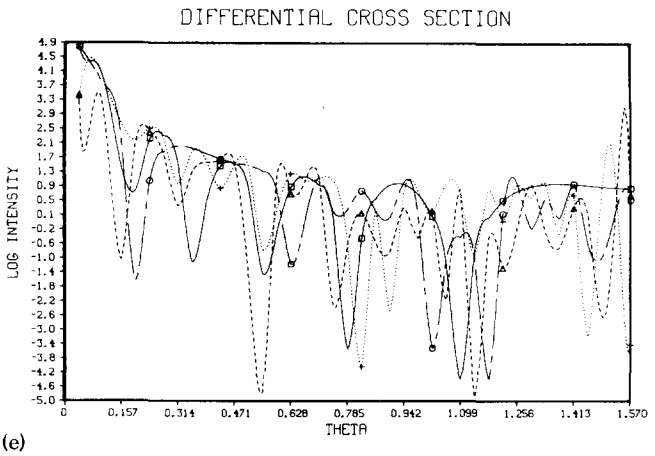
(b)



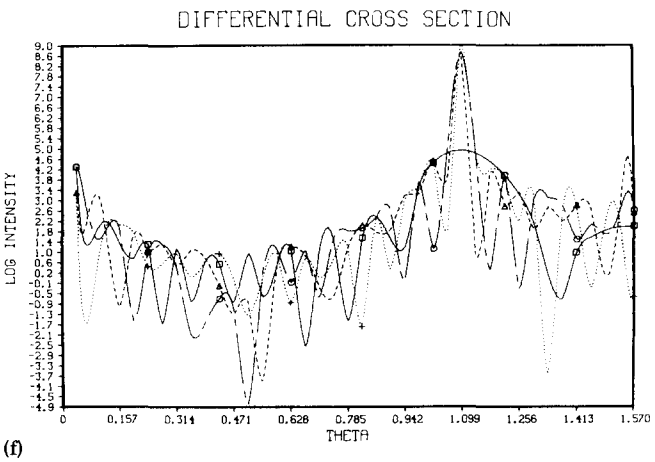
(c)



(d)



(e)



(f)

FIG. 8. Plots of $\log I(\theta, \pi/4)$ versus scattering angle θ for various values of N, N being the number of atoms along each of the axes of the crystal. The wavefunction of the surface scatters corresponds to the one bound state $\operatorname{sech}^2\gamma x$ potential. In each of the graphs the incident particle energy k_0^2, the energy state of the surface valence electron, and the ratio of the range of the surface potential to lattice bond length l'/l are fixed. (a) $l'/l = 1.5$, $k_0^2 = 5$; (b) $l'/l = 1.5$, $k_0^2 = 10$; (c) $l'/l = 15$, $k_0^2 = 5$; (d) $l'/l = 15$, $k_0^2 = 10$; (e) $l'/l = \infty$, $k_0^2 = 5$; (f) $l'/l = \infty$, $k_0^2 = 10$.

Legend: Number of surface atoms along axes (N) and corresponding surface state energies, $mE/\hbar^2\gamma^2 = k^2/2\gamma^2$.

+ -----10 000 atoms,	$k^2/2\gamma^2 = 0.041\,989$;
Δ -----1 000 atoms,	$k^2/2\gamma^2 = 0.041\,989$;
○ -----100 atoms,	$k^2/2\gamma^2 = 0.042\,015$;
□ -----10 atoms,	$k^2/2\gamma^2 = 0.044\,674$.

In Figs. 8 we have plotted the $\log I(\theta, \pi/4)$ versus scattering angle θ for a number of values of N. In each graph the incident particle energy k_0^2, the energy state of the surface valence electron, and the ratio of the range of the surface potential to lattice bond length l'/l are fixed. These plots indicate that the overall intensity profile is not greatly affected by the l'/l ratio, but is clearly dependent on the total number of atoms on the surface. Combining this information with that of Fig. 9 which depicts the variation of $\log I(\theta, \pi/4)$ versus incident particle energy, leads one to believe that for a fixed surface state energy one need only require a sur-

face of $\sim 100^2$ atoms to reproduce for particular values of k_0^2 and θ sharp Bragg diffraction lines. By assuming an interatomic lattice spacing of 1 Å, one can obtain a proper crystallographic representation of a material from a sample having an area of only 10^4 Å^2. This figure lies within the experimental accuracy of ELEED.

One can also discern, in qualitative agreement with experiment, the secondary fine structure which increases in intensity in the vicinity of the Bragg angles.

The total scattering cross section above the surface may be found by integrating $I(\theta,\phi)$ over all angles:

$$\sigma(k_0) = |B|^2 \int_0^{2\pi}\int_0^{\pi/2} d\phi\, d\theta \sin\theta \frac{\sin^2((Nk_0/2)\sin\theta\cos\phi)\sin^2((Nk_0/2)\sin\theta\sin\phi)}{\sin^2((k_0/2)\sin\theta\cos\phi)\sin^2((k_0/2)\sin\theta\sin\phi)} . \tag{6.17}$$

By employing the identities

$$2\left(\frac{N}{2} + \sum_{m=1}^{N-1}(N-m)\cos mx\right) = \frac{\sin^2(Nx/2)}{\sin^2(x/2)} , \tag{6.18a}$$

$$\cos A\cos B = \tfrac{1}{2}[\cos(A+B)+\cos(A-B)], \tag{6.18b}$$

$$\int_0^{2\pi}\cos(c\cos\phi)\,d\phi = \int_0^{2\pi}\cos(c\sin\phi)\,d\phi = 2\pi J_0(c), \tag{6.18c}$$

$$\int_0^{2\pi}\cos[c(a\cos\phi + b\sin\phi)]\,d\phi = 2\pi I_0(ic\sqrt{a^2+b^2}) = 2\pi J_0(c\sqrt{a^2+b^2}), \tag{6.18d}$$

$$\int_0^{\pi/2} d\theta \sin\theta J_0(c'\sin\theta) = \sqrt{\pi/2c'}\,J_{1/2}(c') = \mathrm{sinc}\,c'/c', \tag{6.18e}$$

(6.17) becomes

$$\sigma(k_0) = 2\pi|B|^2\left(N^2 + 4N\sum_{m=1}^{N-1} j_0(mk_0)(N-m) + 4\sum_{m,l=1}^{N-1}(N-m)(N-l)j_0(k_0\sqrt{m^2+l^2})\right). \tag{6.19}$$

Thus, the total cross section is a slowly oscillating function having a maximum value of $2\pi|B|^2N^4$ at $k_0 = 0$ and reaches a minimum value of $2\pi|B|^2N^2$ as $k_0 \to \infty$. Classically, these results would be four times smaller. It should be noted that due to the imposition of the condition that the scattering phenomenon takes place only above the crystal surface $\sigma(k_0)$ is one-half the value obtained for quantum mechanical coherent scattering by point sources.

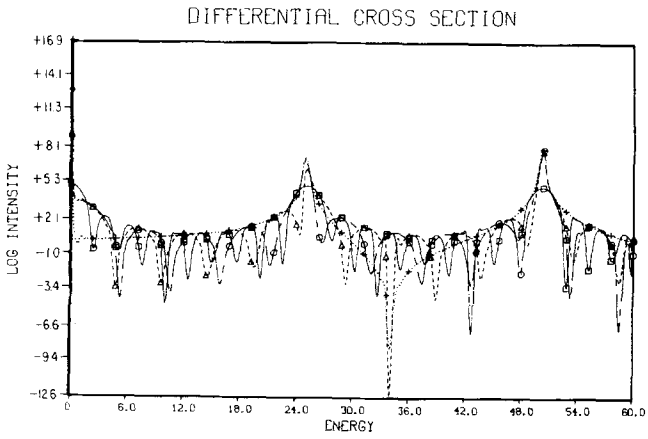


FIG. 9. Variation of $\log I(\theta,\pi/4)$ versus incident particle energy k_0^2 when $l'/l = \infty$. All the other conditions of Fig. 8 are upheld (cf. Fig. 1b).

The above formalism can be extended to the study of nonorthogonal lattices, i.e., simple triclinic. Now, the positions of the surface atoms may be denoted by the vector $\rho(j_1,j_2) = (j_1\hat{e}_1, j_2\hat{e}_2, 0\hat{e}_3)$, with the $\hat{e}_i$ corresponding to the three nonorthogonal unit vectors. By orienting the crystal so that $\hat{e} \equiv \underset{\sim}{i}$, $\hat{e}_2$ may be written as

$$\hat{e}_2 = -\underset{\sim}{i}\sin\gamma + \underset{\sim}{j}\cos\gamma, \tag{6.20}$$

γ being the angle between $\hat{e}_2$ and $\underset{\sim}{j}$. We now have a Cartesian representation for $\rho(j_1,j_2)$ and may proceed as before.

7. ELEED FROM B.C.C. CRYSTAL SURFACES

We now consider the more realistic b.c.c. surface scattering problem.

The lattice Green's function analog to equations (3.14) for a b.c.c. lattice with a surface is

$$8F(E)\psi(j_1,j_2,j_3) - \psi(j_1\pm1,j_2+1,j_3+1) - \psi(j_1\pm1,j_2-1,j_3+1) - \psi(j_1\pm1,j_2+1,j_3-1) - \psi(j_1\pm1,j_2-1,j_3-1) = H(j_1,j_2,j_3) \tag{7.1a}$$

where

$$H(j_1,j_2,j_3) = [\delta_{j_3,N_3} + \delta_{j_3,1}](4\eta)\psi(j_1,j_2,j_3) - \delta_{j_3,1}[\psi(j_1\pm1,j_2+1,N_3) + \psi(j_1\pm1,j_2-1,N_3)] - \delta_{j_3,N_3}[\psi(j_1\pm1,j_2+1,1) + \psi(j_1\pm1,j_2-1,1)]. \tag{7.1b}$$

Substitution of (3.15) into Eqs. (7.1) results in

$$2\hat{F}_{n_1,n_2}\phi(j_3) - \phi(j_3\pm1) = \delta_{j_3,N_3}[\hat{\eta}\phi(N_3) - \phi(1)] + \delta_{j_3,1}[\hat{\eta}\phi(1) - \phi(N_3)] \tag{7.2}$$

where we define

$$\hat{F}_{n_1,n_2} = \frac{F(E)}{\cos(2\pi n_1/N_1)\cos(2\pi n_2/N_2)} \tag{7.3a}$$

and

$$\hat{\eta} = \frac{\eta}{\cos(2\pi n_1/N_1)\cos(2\pi n_2/N_2)} . \tag{7.3b}$$

This formalism is now identical to that found for the s.c.c. lattice. The same scattering formulas can now be used, after replacement of all F_{n_1,n_2} and η functions by $\hat{F}_{n_1,n_2}$ and $\hat{\eta}$. As in the previous sections, all results are analytic. The scattering intensities will resemble the s.c.c. intensities, but peaks will be shifted.

8. CONCLUSION

A method has been introduced which uses the conservative property of current on a network in order to characterize the scattering of charged particles from a finite number of defects in a periodic lattice of atoms. Since the current conservation is a topological property of the network, the approximations inherent to the model are entirely physical. For calculational ease, we have used a one electron approximation to a wavefunction which is the solution of the Schrödinger equation when the atomic potential $V(x) = (-\gamma^2\hbar^2/m)\,\mathrm{sech}^2\gamma x$. The use of this potential is convenient since as $\gamma \to 0$ we obtain the Sommerfeld free electron model and as $\gamma \to \infty$ we obtain the tight binding model. The parameter γ can therefore be likened to an effective mass and thereby enables one to fit calculations to experiment. The results of this method can also be put into a one to one correspondence with the more established tight binding model[15] with the identification of the Hamiltonian parameters $\alpha^{(m)}$ (the orbital energy of the mth orbital) and $\beta^{(m)}_{(n)}$ (the hopping matrix element between the mth and nth orbitals) to functions arising from (2. 6) and its derivative. Thus, for nearest neighbor interactions between orbitals

$$\beta^{(m)}_{(m\pm1)} = \frac{-k(1+\gamma^2/k^2)}{\sin kl + (\gamma/k)\cos kl\tanh\gamma l}, \tag{8.1}$$

$$\alpha^{(m)} = \frac{-k(\cos kl + \gamma/k^2[k\cos kl\,\mathrm{sech}^2\gamma l - k\sin kl\tanh\gamma l])}{\sin kl + (\gamma/k)\cos kl\tanh\gamma l}.$$

One should note, however, that this model only has one arbitrary parameter as compared with the two in the tight binding approximation. This lends itself to a certain amount of simplification in fitting theoretical results with experiment. Since qualitative agreement with experiment is excellent, this seems to imply that structure dominates the electronic character of a scattering system. That structure is not the whole story is aptly brought out by Duke[1] in his discussion of surface scattering. Thus, an obvious and necessary extension of this approach is the inclusion of many body effects. As a first approximation, however, one can utilize this type of method to get a grip on what to expect from a much more complicated and certainly an orders of magnitude more costly approach. Work is now being completed in extending this procedure to the multiple scattering formalism of Lloyd and Keller, where now the potential field of the atom for the valence electron is an experimentally fitted 3D potential and the scattering occurs along the channels of the scatterer, which now corresponds to the network of our system.

ACKNOWLEDGMENTS

The author would like to thank Dr. U. Cattaneo, Dr. U. Landman, and Professor J. Weare for their helpful comments. He also wishes to thank Professor E. W. Montroll for supplying the financial wherewithal during the course of this work.

APPENDIX A[16,17]

Let M_0 be an N-dimensional cyclic matrix of the form

$$M_0 = \begin{pmatrix} m(0) & m(1) & \cdots & m(N-1) \\ m(-1) & m(0) & m(1) & \cdots \\ \vdots & m(-1) & m(0) & \cdots \\ m(1) & \cdots & \cdots & m(0) \end{pmatrix}$$

$$= \sum_{j=0}^{N-1} m(j)R^j \tag{A1}$$

where the permutation matrix R is defined by

$$R = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & \\ 0 & 0 & 0 & 1 & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad R^2 = \begin{pmatrix} 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ \vdots & & & & \\ 1 & \cdots & & & \\ 0 & 1 & \cdots & & \end{pmatrix} \tag{A2}$$

and

$$R^N = I_N.$$

Since the eigenvalues of R are $\exp(2\pi ik/N)$, $k = 0, \ldots, N-1$, the eigenvalues λ_k of M_0 are as a result of (A1)

$$\lambda_k = \sum_{j=0}^{N-1} m(j)\exp(2\pi ik/N), \tag{A3}$$

If for higher dimensional lattices the elements $m(s)$ are circulant matrices in their own right, the eigenvalue analog to (A3) is

$$\lambda_{k_1,k_2,\ldots,k_n} = \sum_{j_1=0}^{N_1-1} \cdots \sum_{j_n=0}^{N_n-1} m(j_1,\ldots,j_n)\exp\left(2\pi i\sum_{p=1}^{n}\frac{j_pk_p}{N_p}\right) \tag{A4a}$$

with the corresponding normalized eigenvectors

$$\psi_{k_1,\ldots,k_n}(l_1,\ldots,l_n) = \prod_{p=1}^{n} N_p^{-1/2}\exp\left(2\pi i\sum_{p=1}^{n}\frac{l_pk_p}{N_p}\right), \tag{A4b}$$

n being the dimensionality of the lattice.

Consider the eigenvalue matrix equation

$$M_0\psi_{\mathbf{k}} = \lambda_{\mathbf{k}}\psi_{\mathbf{k}} \tag{A5}$$

which when expanded is

$$\lambda_{\mathbf{k}}\psi_{\mathbf{k}}(1) = \sum_{1'} m(1-1')\psi_{\mathbf{k}}(1'). \tag{A6}$$

$m(1-1')$ has the form

$$m(1-1') = \sum_{\mathbf{k}} \lambda_{\mathbf{k}}\psi_{\mathbf{k}}(1)\psi^*_{\mathbf{k}}(1'), \tag{A7a}$$

since if we multiply $m(1-1')$ by $\psi_{\mathbf{k}'}(1')$ and sum over $1'$

$$\sum_{1'} m(1-1')\psi_{\mathbf{k}'}(1') = \sum_{1',\mathbf{k}} \lambda_{\mathbf{k}}\psi_{\mathbf{k}}(1)\psi^*_{\mathbf{k}}(1')\psi_{\mathbf{k}'}(1') \tag{A7b}$$

$$= \sum_{\mathbf{k}'} \lambda_{\mathbf{k}'}\psi_{\mathbf{k}'}(1), \tag{A7c}$$

(A7c) being the result of the orthogonality of the ψ's.

Raising (A7a) to the nth power we find

$$m^n(1-1') = \sum_{\mathbf{k}} \lambda^n_{\mathbf{k}}\psi_{\mathbf{k}}(1)\psi^*_{\mathbf{k}}(1'). \tag{A8}$$

In particular, when $n = -1$ and the elements of M_0 are taken from (3. 3a), the connectivity matrix of a one-dimensional ring, (A8) becomes

$$m^{-1}(l-l') = \frac{1}{2N}\sum_{k=1}^{N}\frac{\exp[2\pi i(l-l')k/N]}{F(E)-\cos(2\pi k/N)}, \tag{A9}$$

the Green's function for a 1D periodic ring. The d-dimensional s.c.c. Green's function can be derived in a similar fashion from (A8) and is equivalent to (3.8) in the main text.

APPENDIX B

The memory function of (3.9) is derivable in the following manner. Consider Figs. 4. Here we depict a one-dimensional ring of N atoms cut between atoms N and 1. From symmetry considerations let us restrict ourselves to the dangling bond connected to atom N. Defining the wavefunction along the dangling bond by $\phi(x)$, $0 \le x \le l'$, then the wavefunction $\psi(N)$, at atom N, is according to (2.6)

$$\psi(N) = \phi(0) = a\cos\delta \tag{B1a}$$

and

$$\phi(x) = \psi(N)\{[\cos kx - (\gamma/k)\sin kx\tanh\gamma x] - \tan\delta_{NE}[\sin kx + (\gamma/k)\cos kx\tanh\gamma x]\}, \tag{B1b}$$

the subscript E of δ_{NE} identifying the end of the chain. If as a boundary condition $\phi(l') = 0$, then

$$\tan\delta_{NE} = \frac{\cos kl' - (\gamma/k)\sin kl'\tanh\gamma l'}{\sin kl' + (\gamma/k)\cos kl'\tanh\gamma l'}. \tag{B2}$$

Since the standard relation between two interior node points [i.e., atoms N and $N-1$ as derived in Ref. 4, Eq. (38a)] is

$$-(1+cu+sv)\psi(N-1) = [2s(c+u)\tan\delta_{N,N-1} + s^2 - c^2 + sv - cu]\psi(N), \tag{B3}$$

then by applying the conservation condition (2.2) we obtain

$$0 = \tan\delta_{NE} + \tan\delta_{N,N-1} \tag{B4}$$

which after some algebraic rearrangement becomes

$$2F(k/\gamma)\psi(N) - \psi(N-1) - \psi(1) = \eta\psi(N) - \psi(1) \tag{B5}$$

where

$$\eta = F(k/\gamma) - \frac{2s(c+u)}{1+cu+sv}\times\tan\delta_{NE} \tag{B6a}$$

and

$$F(k/\gamma) = (c^2 - s^2 + cu - sv)/(1+cu+sv). \tag{B6b}$$

*Work partially supported by Grant No. GU-4040 of the National Science Foundation.

†Present address: Department of Chemistry, University of California, La Jolla, California 92037.

[1] C.B. Duke, *The Characterization of Solid Surfaces; Elastic Low Energy Electron Diffraction; Surface Crystallography.*
[2] C.J. Davisson and L.H. Germer, Phys. Rev. **30**, 705 (1927).
[3] E.W. Montroll, J. Math. Phys. **11**, 635 (1970).
[4] R.G. Mills and E.W. Montroll, J. Math. Phys. **11**, 2525 (1970).
[5] E.W. Montroll, J. Phys. Chem. Solids **34**, 597 (1973).
[6] J. Keller and P.V. Smith, J. Phys. C **5**, 1109 (1972).
[7] A.G. Sitenko, *Lectures in Scattering Theory* (Pergamon, Oxford, 1971).
[8] G.F. Koster, Phys. Rev. **95**, 1436 (1954).
[9] A.A. Maradudin, E.W. Montroll, G.H. Weiss, and I.P. Ipatova, *Theory of Lattice Dynamics in the Harmonic Approximation, Solid State Physics, Supplement* 3 (Academic, New York, 1971), p. 377.
[10] C. Berge and A. Ghouila-Houri, *Programming, Games, and Transportation Networks* (Methuen, London, 1965).
[11] K. Ruedenberg and C.W. Scherr, J. Chem. Phys. **21**, 1565 (1953).
[12] A.B. Budgor, Ph.D. thesis, University of Rochester, 1974.
[13] C. Kittel, *Introduction to Solid State Physics* (Wiley, New York, 1971), 4th ed., p. 78.
[14] M. Born and E. Wolf, *Principles of Optics* (Pergamon, New York, 1964), pp. 392–410.
[15] J.S. Dancz and S.F. Edwards, J. Phys. C **6**, 3413 (1973).
[16] P.B. Abraham, *Calculation of Functionals of Matrices Arising in Solid State Physics and Quantum Chemistry* (Goddard Space Flight Center, 1966), pp. 197–199, 222.
[17] E.W. Montroll, Lectures on Network Models in Solid State Physics, University of Rochester, 1969.

Positive solution of a time and energy dependent neutron transport problem

C. V. Pao

Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27607
(Received 27 September 1974; revised manuscript received 20 January 1975)

The purpose of this paper is to give a constructive method for the determination of a solution and an existence–uniqueness theorem for some nonlinear time and energy dependent neutron transport problems, including the linear transport system. The geometry of the medium under consideration is allowed to be either bounded or unbounded which includes the geometry of a finite or infinite cylinder, a half-space and the whole space R^m ($m = 1,2,\cdots$). Our approach to the problem is by successive approximation which leads to various recursion formulas for the approximations in terms of explicit integrations. It is shown under some Lipschitz conditions on the nonlinear functions, which describe the process of neutrons absorption, fission, and scattering, that the sequence of approximations converges to a unique positive solution. Since these conditions are satisfied by the linear transport equation, all the results for the nonlinear system are valid for the linear transport problem. In the general nonlinear problem, the existence of both local and global solutions are discussed, and an iterative process for the construction of the solution is given.

1. INTRODUCTION

In the neutron transport and radiative transfer problems, the usual time and energy dependent equation for the density function may be written as the following integro-differential equation (usually referred to as linearized Boltzmann equation):

$$\frac{\partial N}{\partial t} + v(\Omega\cdot\nabla N + \Sigma(t,x,\Omega,\mu)N) = \int_0^{\mu_m}\int_S \Sigma_s(t,x,\Omega,\Omega',\mu,\mu')N(t,x,\Omega',\mu')\,d\Omega'\,d\mu' + q(t,x,\Omega,\mu) \quad (t>0,\ x\in D,\ \Omega\in S,\ 0\le\mu\le\mu_m), \tag{1.1}$$

where D is an open domain in the Euclidean space R^m ($m=1,2,\cdots$), S is the unit sphere in R^m, and ∇ is the gradient operator with respect to the spatial variable $x\in R^m$. In terms of neutron transport, $N=N(t,x,\Omega,\mu)$ is the neutron density at time t, position x moving in the direction Ω with speed v energy μ, μ_m is the maximum energy, Σ, Σ_s are the total and scattering cross sections, respectively, and q is the neutron source. The equation in the form of (1.1) has been investigated by many workers and a number of methods have been developed for the determination of the solution (cf. Refs. 1—16). Following a classical approach for initial boundary value problem by Fourier transformation, Case and Zweifel[1,2] transformed a linear time-dependent equation into a time-independent equation and then investigated the transformed problem through the construction of a Green's function. A different approach was developed by Bellman, Kalaba, and Wing[3,4] under the title of Invariant Imbedding. This approach reduces the linear transport problem into a nonlinear initial value problem which has its computational significance. Using the linear semigroup theory, Lehner and Wing[5,6] investigated the linear transport problem in a slab geometry by studying the spectral properties of the linear transport operator. Their approach was extended to spatial domains of higher dimensions by Hejtmanek[8], Suhadolc[14], Vidav[15], Mika[16], and many others and, to some extent, by Reed[13] for a moving body. More recently, this investigator[10,11] treated some nonlinear transport problems by the method of successive approximations. An essential aspect of this method is that it leads to a recursion formula for the determination of the solution. Nevertheless, most of the work mentioned above concerned with energy-independent transport problems in bounded spatial domains. In this paper, we treat a nonlinear time and energy dependent transport equation, including the linear equation (1.1), in an arbitrary medium (either bounded or unbounded). Specifically, we consider the following nonlinear Boltzmann-type equation:

$$\frac{\partial N}{\partial t} + v(\Omega\cdot\nabla N + \sigma(t,x,\Omega,\mu,N)) = \int_0^{\mu_m}\int_S \sigma_s(t,x,\Omega,\Omega',\mu,\mu',N(t,x,\Omega',\mu'))\,d\Omega'\,d\mu' + q(t,x,\Omega,\mu) \quad (t>0,\ x\in D,\ \Omega\in S,\ 0\le\mu\le\mu_m). \tag{1.2}$$

The consideration of a nonlinear equation includes the effect of collisions among neutrons where σ, σ_s are, in general, nonlinear in N. The spatial domain D is determined by the geometry of the transport medium under consideration. Here we consider an arbitrary convex geometry in the sense that D can be either bounded or unbounded. The unboundedness of D includes the most commonly discussed geometries such as the whole space R^m, a half-space, an infinite or semi-infinite cylinder, etc. The velocity $v\Omega$ is assumed to range over a bounded set in R^m.

In addition to Eq. (1.2) we consider the following boundary condition:

$$N(t,x,\Omega,\mu)=0 \quad (t>0,\ x\in\Gamma,\ \Omega \text{ incoming},\ 0\le\mu\le\mu_m),$$
$$\lim_{|x|\to\infty} N(t,x,\Omega,\mu)=0 \quad (t>0,\ \Omega\in S,\ 0\le\mu\le\mu_m), \tag{1.3}$$

and the initial condition

$$N(0, x, \Omega, \mu) = \phi(x, \Omega, \mu) \quad (x \in D,\ \Omega \in S,\ 0 \leq \mu \leq \mu_m), \tag{1.4}$$

where Γ is the boundary of D. The physical meaning of the boundary condition (1.3) is that no neutrons enter the medium through the boundary surface Γ and neutrons vanish as $|x|$ approaches infinite. It is to be pointed out that the boundary Γ is allowed to be either empty or the whole boundary of D (that is, either $D = R^m$ or D is bounded). In this situation, only one of the conditions in (1.2) appears.

The purpose of this paper is twofold: (i) To derive various recursion formulas for the calculation of approximate solutions and error estimates for the approximations and (ii) to give an existence—uniqueness theorem by proving the convergence of the approximations to a unique positive solution of the problem. We also discuss the existence of a local solution and its continuation to a global solution. This result leads to an interesting analogy between the initial boundary-value problem (1.2)—(1.4) and the calssical existence—uniqueness theorem of the Cauchy problem for ordinary differential equations. All the results for the nonlinear problem are applied to the linear problem (1.2), (1.3), (1.4). Since each formula given in this paper involves only successive integration of a known function, our method yields both analytical results and computational significance.

2. EXISTENCE PROBLEM BY SUCCESSIVE APPROXIMATIONS

Throughout the paper we always assume that the functions σ, σ_s, q, ϕ are bounded continuous for $(t, x, \Omega, \mu) \in E$, $|N| < \infty$, $q(t, x, \Omega, \mu) \to 0$, $\phi(x, \Omega, \mu) \to 0$ uniformly in (t, Ω, μ) as $|x| \to \infty$ and ϕ satisfies the boundary condition (1.3) at $t = 0$, where $E = [0, T] \times D \times S \times [0, \mu_m]$ and T is an arbitrary finite number. It is assumed, for convenience, that $v = 1$ (see Remark 2.1). In addition, we make the following assumptions:

(H): $\sigma(t, x, \Omega, \mu, 0) = \sigma_s(t, x, \Omega, \Omega', \mu, \mu', 0) = 0$ and there exist bounded nonnegative functions ρ, ρ_s satisfying

$$\bar{\rho} = \sup_{(t,x,\Omega,\mu) \in E} [\rho(t, x, \Omega, \mu) + \int_0^{\mu_m} \int_S \rho_s(t, x, \Omega, \Omega', \mu, \mu')\, d\Omega'\, d\mu'] < \infty \tag{2.1}$$

such that for $t \in [0, T]$, $x \in D$, $\Omega, \Omega' \in S$, $\mu, \mu' \in [0, \mu_m]$, $\eta_1, \eta_2 \in (-\infty, \infty)$,

$$|\sigma(t, x, \Omega, \mu, \eta_1) - \sigma(t, x, \Omega, \mu, \eta_2)| \leq \rho(t, x, \Omega, \mu)\, |\eta_1 - \eta_2|, \tag{2.2}$$

$$|\sigma_s(t, x, \Omega, \Omega', \mu, \mu', \eta_1) - \sigma_s(t, x, \Omega, \Omega', \mu, \mu', \eta_2)| \leq \rho_3(t, x, \Omega, \Omega', \mu, \mu')\, |\eta_1 - \eta_2|. \tag{2.3}$$

The above hypothesis implies that if N is a bounded continuous function on E such that $N(t, x, \Omega, \mu) \to 0$ uniformly in (t, Ω, μ) as $|x| \to \infty$ then so is the function

$$f_0(t, x, \Omega, \mu, N) \equiv -\sigma(t, x, \Omega, \mu, N) + \int_0^{\mu_m} \int_S \sigma_s(t, x, \Omega, \Omega', \mu, \mu', N(t, x, \Omega', \mu'))\, d\Omega'\, d\mu' + q(t, x, \Omega, \mu). \tag{2.4}$$

Notice that the uniform convergence of f_0 to zero as $|x| \to \infty$ follows from (2.2), (2.3) with $\eta_2 = 0$.

In order to describe our process of successive approximation and to insure the convergence of the approximations to a solution, we first transform the problem (1.2)—(1.4) by letting $u = \exp(-\lambda t)N$ to obtain the transformed system

$$(\partial/\partial t + \Omega \cdot \nabla)u + \lambda u = f(t, x, \Omega, \mu, u) \quad [(t, x, \Omega, \mu) \in E], \tag{2.5}$$

$$u(t, x, \Omega, \mu) = 0 \quad (t \in (0, T],\ x \in \Gamma,\ \Omega \text{ incoming}, \ 0 \leq \mu \leq \mu_m), \tag{2.6}$$

$$\lim_{|x| \to \infty} u(t, x, \Omega, \mu) = 0 \quad (t \in (0, T],\ \Omega \in S,\ 0 \leq \mu \leq \mu_m),$$

$$u(0, x, \Omega, \mu) = \phi(x, \Omega, \mu) \ (x \in D,\ \Omega \in S,\ 0 \leq \mu \leq \mu_m), \tag{2.7}$$

where $\lambda > \bar{\rho}$ is a constant and

$$f(t, x, \Omega, \mu, u) \equiv \exp(-\lambda t) f_0(t, x, \Omega, \mu, \exp(\lambda t)u). \tag{2.8}$$

We then seek a solution for the linear equation

$$(\partial/\partial t + \Omega \cdot \nabla)u + \lambda u = h \quad [(t, x, \Omega, \mu) \in E] \tag{2.9}$$

under the boundary and initial conditions (2.6), (2.7), where h is a bounded continuous function on E such that $h(t, x, \Omega, \mu) \to 0$ uniformly in (t, Ω, μ) as $|x| \to \infty$. By using the new variables $x' = x - \Omega t$, $\tau = t$, Eq. (2.9) reduces to

$$\frac{\partial u}{\partial \tau}(\tau, x' + \Omega\tau, \Omega, \mu) + \lambda u(\tau, x' + \Omega\tau, \Omega, \mu) = h(\tau, x' + \Omega\tau, \Omega, \mu), \tag{2.10}$$

where $\partial u/\partial \tau$ is the substantial derivative of u. Multiplication by $\exp(\lambda\tau)$ and integration from 0 to t over τ lead to

$$\exp(\lambda t)u(t, x' + \Omega t, \Omega, \mu) - u(0, x', \Omega, \mu) = \int_0^t \exp(\lambda\tau) h(\tau, x' + \Omega\tau, \Omega, \mu)\, d\tau. \tag{2.11}$$

Applying the initial condition (2.7) and replacing x' by $x - \Omega t$, we obtain

$$u(t, x, \Omega, \mu) = \exp(-\lambda t)\, \phi(x - \Omega t, \Omega, \mu) + \int_0^t \exp[-\lambda(t - \tau)]\, h(\tau, x - \Omega(t - \tau), \Omega, \mu)\, d\tau. \tag{2.12}$$

It is clear by reversing the order in the above derivation that the function u given by (2.12) satisfies (2.10) and therefore is a solution of (2.9) when $(\partial/\partial t + \Omega \cdot \nabla)u$ is considered as the substantial derivative of u. Obviously, u also satisfies the initial condition (2.7). To show that u satisfies the boundary condition (2.6), we define

$$\phi(x, \Omega, \mu) = h(t, x, \Omega, \mu) = 0 \quad \text{if } x \notin \bar{D} \text{ and } \Omega \text{ incoming}, \tag{2.13}$$

where $\bar{D}$ is the closure of D. Then u satisfies the first condition in (2.6) since for $x \in \Gamma$ and incoming Ω, the point $x - \Omega(t-\tau)$ is outside of D for every $t > \tau \geq 0$. In view of the uniform convergence to zero of the functions h and ϕ as $|x| \to \infty$, we see that $u(t, x, \Omega, \mu) \to 0$ uniformly in (t, Ω, μ) as $|x| \to \infty$. Hence the function u given by (2.10) is the solution of the problem (2.9), (2.6), (2.7), where $(\partial/\partial t + \Omega \cdot \nabla)u$ is considered as the substantial derivative of u. Notice that if h is differentiable in each component x_i of x, then u is differentiable in t and x_i and thus is a solution of (2.9), (2.6), (2.7) in the usual sense.

The above observation leads to the following consideration: Define

$$\sigma(t, x, \Omega, \cdot) = \sigma_s(t, x, \Omega, \cdot) = q(t, x, \Omega, \mu) = \phi(x, \Omega, \mu) = 0 \quad \text{when } x \notin \bar{D},\ \Omega \text{ incoming} \tag{2.14}$$

and

$$(f(u))(t, x, \Omega, \mu) = f(t, x, \Omega, \mu, u(t, x, \Omega, \mu)). \tag{2.15}$$

Since $f(u)$ is bounded continuous on E such that $(f(u))(t, x, \Omega, \mu) \to 0$ uniformly as $|x| \to \infty$ whenever u has this property and since by (2.14), $(f(u))(t, x, \Omega, \mu) = 0$ when $x \notin \bar{D}$, we see by letting $h = f(u)$ in (2.12) that the problem (2.5)—(2.7) is reduced to the integral equation

$$u(t, x, \Omega, \mu) = \exp(-\lambda t)\,\phi(x - \Omega t, \Omega, \mu) + \int_0^t \exp[-\lambda(t-\tau)] \times (f(u))(\tau, x - \Omega(t-\tau), \Omega, \mu)\, d\tau \quad [(t, x, \Omega, \mu) \in E]. \tag{2.16}$$

Hence if we can find a solution u of the above integral equation and if u is bounded continuous on E such that $u(t, x, \Omega, \mu) \to 0$ uniformly as $|x| \to \infty$, then u is a solution of the problem (2.5)—(2.7) along the characteristics of the streaming operator. We shall accomplish this by the method of successive approximation. Throughout the paper, by a solution of (2.5)—(2.7) we mean a bounded continuous function u which satisfies the conditions (2.6), (2.7) and Eq. (2.5), where $(\partial/\partial t + \Omega \cdot \nabla)u$ is considered as the substantial derivative of u. A similar definition holds for a solution of the problem (1.2)—(1.4).

Let $C(E)$ be the Banach space of all bounded continuous functions u on E equipped with the norm

$$\|u\| = \sup\{|u(t, x, \Omega, \mu)|;\ (t, x, \Omega, \mu) \in E\}$$

and let $C_0(E)$ be the closed subspace of all functions $u \in C(E)$ such that $u(t, x, \Omega, \mu) \to 0$ uniformly in (t, Ω, μ) as $|x| \to \infty$. Define an operator $\mathcal{A}$ on $C_0(E)$ by

$$(\mathcal{A}u)(t, x, \Omega, \mu) = \exp(-\lambda t)\,\phi(x - \Omega t, \Omega, \mu) + \int_0^t \exp[-\lambda(t-\tau)] \times (f(u))(\tau, x - \Omega(t-\tau), \Omega, \mu)\, d\tau. \tag{2.17}$$

In view of the hypothesis H, $\mathcal{A}$ maps $C_0(E)$ into itself. Hence Eq. (2.16) becomes the operator equation

$$u = \mathcal{A}u \quad [u \in C_0(E)] \tag{2.18}$$

in the space $C_0(E)$. Let $u^{(0)} \in C_0(E)$ be given. Then we can determine a sequence $\{u^{(k)}\}$ successively from

$$u^{(k)} = \mathcal{A}u^{(k-1)}, \quad k = 1, 2, \cdots. \tag{2.19}$$

To insure that this sequence converges to a solution of (2.18), we show the following:

Lemma 2.1: $\mathcal{A}$ is a contraction mapping on $C_0(E)$ with a contraction constant $(\bar{\rho}/\lambda)$.

Proof: Let $u_1, u_2 \in C_0(E)$. Then by (2.17),

$$\begin{aligned} &|(\mathcal{A}u_1 - \mathcal{A}u_2)(t, x, \Omega, \mu)| \\ &\leq \int_0^t \exp[-\lambda(t-\tau)]\, |(f(u_1) - f(u_2))(\tau, x - \Omega(t-\tau), \Omega, \mu)|\, d\tau, \\ &\leq \left(\int_0^t \exp[-\lambda(t-\tau)]\, d\tau\right) \|f(u_1) - f(u_2)\| < \lambda^{-1}\|f(u_1) - f(u_2)\| \end{aligned}$$

for any $(t, x, \Omega, \mu) \in E$. Hence

$$\|\mathcal{A}u_1 - \mathcal{A}u_2\| \leq \lambda^{-1}\|f(u_1) - f(u_2)\| \quad [u_1, u_2 \in C_0(E)]. \tag{2.20}$$

Since, by (H),

$$\begin{aligned} |(f(u_1) - f(u_2))(t, x, \Omega, \mu)| \leq \exp(-\lambda t)\Big(&\rho(t, x, \Omega, \mu)\, | \\ &\times |\exp(\lambda t)u_1 - \exp(\lambda t)u_2| \\ &+ \int_0^{\mu_m}\int_S \rho_s(t, x, \Omega, \Omega', \mu, \mu')\, |\exp(\lambda t)\, u_1 \\ &- \exp(\lambda t)\, u_2|\, d\Omega' d\mu'\Big) \\ \leq \bar{\rho}\|u_1 - u_2\| \end{aligned}$$

for $(t, x, \Omega, \mu) \in E$, we see that

$$\|f(u_1) - f(u_2)\| \leq \bar{\rho}\|u_1 - u_2\| \quad [u_1, u_2 \in C_0(E)]. \tag{2.21}$$

It follows from (2.20), (2.21) that

$$\|\mathcal{A}u_1 - \mathcal{A}u_2\| \leq (\bar{\rho}/\lambda)\|u_1 - u_2\| \quad [u_1, u_2 \in C_0(E)]. \tag{2.22}$$

By the choice of $\lambda > \bar{\rho}$, $\mathcal{A}$ is a contraction mapping on $C_0(E)$ with a contraction constant $(\bar{\rho}/\lambda) < 1$. This proves the lemma.

The above lemma leads immediately to the following conclusion.

Theorem 2.1: Assume that (H) holds. Then for any $u^{(0)} \in C_0(E)$ the sequence $\{u^{(k)}\}$ given successively by

$$u^{(k)}(t, x, \Omega, \mu) = \exp(-\lambda t)\,\phi(x - \Omega t, \Omega, \mu) + \int_0^t \exp[-\lambda(t-\tau)] \times (f(u^{(k-1)}))(\tau, x - \Omega(t-\tau), \Omega, \mu)\, d\tau, \quad k = 1, 2, \cdots, \tag{2.23}$$

converges in $C_0(E)$ to a unique solution $\tilde{u}$ of the integral equation (2.16). Moreover, $\tilde{u}$ is a solution of (2.5)—(2.7) along the characteristics of the streaming operator and the approximations $u^{(k)}$ satisfy the error estimates

$$\|u^{(k)} - \tilde{u}\| \leq [\bar{\rho}/(\lambda - \bar{\rho})]\,(\bar{\rho}/\lambda)^{(k-1)}\,\|u^{(1)} - u^{(0)}\|, \quad k = 1, 2, \cdots. \tag{2.24}$$

Proof: By Lemma 2.1 and the contraction mapping theorem the sequence $\{u^{(k)}\}$ determined from (2.19) converges to a unique solution $\tilde{u} \in C_0(E)$ of (2.18). Since (2.18) and (2.19) are the operator equations of (2.16)

and (2.23), respectively, we conclude that the sequence $\{u^{(k)}\}$ given by (2.23) converges in $C_0(E)$ to the unique solution $\tilde{u}$ of (2.16). Since $\tilde{u} \in C_0(E)$ implies $f(\tilde{u}) \in C_0(E)$, we see that $\tilde{u}$ is a solution of (2.5)–(2.7). Finally, the error estimate (2.24) follows immediately from the contraction property of $\mathcal{A}$.

If we multiply (2.23) by $\exp(\lambda t)$ and let $N^{(k)} = \exp(\lambda t)\, u^{(k)}$, $N = \exp(\lambda t)u$, then the problem (2.5)–(2.7) becomes the original problem (1.2)–(1.4) and the recursion formula (2.23) is reduced to

$$N^{(k)}(t,x,\Omega,\mu) = \phi(x-\Omega t,\Omega,\mu) + \int_0^t f_0(\tau, x-\Omega(t-\tau),\Omega,\mu, N^{(k-1)}(\tau, x-\Omega(t-\tau),\Omega,\mu))\,d\tau, \quad k=1,2,\cdots. \tag{2.25}$$

It follows from the convergence of $\{u^{(k)}\}$ to u that the sequence $\{N^{(k)}\}$ given above converges to a unique solution $\tilde{N} = \exp(\lambda t)\tilde{u}$ of the original problem (1.2)–(1.4). This observation leads to the following:

Theorem 2.2: Assume that (H) holds. Then for any $N^{(0)} \in C_0(E)$ the sequence $\{N^{(k)}\}$ given by (2.25) converges in $C_0(E)$ to a unique solution $\tilde{N} \in C_0(E)$ of the problem (1.2)–(1.4).

Remark 2.1: (a) In obtaining the sequence $\{N^{(k)}\}$ given by (2.25) we have assumed that $v=1$. When $v \neq 1$, this sequence should be replaced by

$$N^{(k)}(t,x,\Omega,\mu) = \phi(x-v\Omega t,\Omega,\mu) + \int_0^t \tilde{f}_0(\tau, x-v\Omega(t-\tau),\Omega,\mu, N^{(k-1)}(\tau, x-v\Omega(t-\tau),\Omega,\mu))\,d\tau, \quad k=1,2,\cdots, \tag{2.26}$$

where $\tilde{f}_0$ is defined by (2.4) with σ replaced by $v\sigma$. The above recursion formula can be obtained from (2.9), (2.12) (with Ω replaced by $v\Omega$).

(b) When $\sigma = \Sigma N$, where Σ is a bounded continuous function on E, the sequence $\{N^{(k)}\}$ given by (2.25) converges to a unique solution $\tilde{N}$ of (1.2)–(1.4) (with $v=1$) if σ_s satisfies the condition (2.3). On the other hand, if we define a sequence from the equation

$$\left(\frac{\partial}{\partial t} + \Omega\cdot\nabla\right) N^{(k)} + \Sigma N^{(k)} = g(t,x,\Omega,\mu, N^{(k-1)}(t,x,\Omega,\mu)), \quad k=1,2,\cdots, \tag{2.27}$$

together with the conditions (1.3), (1.4) for $N^{(k)}$, where

$$g(t,x,\Omega,\mu,N) = \int_0^{\mu_m}\int_S \sigma_s(t,x,\Omega,\Omega',\mu,\mu', N(t,x,\Omega',\mu'))\times d\Omega'\,d\mu' + q(t,x,\Omega,\mu), \tag{2.28}$$

then we can solve this system to obtain another recursion formula which is given by

$$N^{(k)}(t,x,\Omega,\mu) = \gamma(t,0,x,\Omega,\mu)\phi(x-\Omega t,\Omega,\mu) + \int_0^t \gamma(t,\tau,x,\Omega,\mu) \times g(\tau, x-\Omega(t-\tau),\Omega,\mu, N^{(k-1)}(\tau, x-\Omega(t-\tau),\Omega,\mu))\,d\tau, \quad k=1,2,\cdots. \tag{2.29}$$

Here γ is defined by

$$\gamma(t,\tau,x,\Omega,\mu) = \exp\left(\int_\tau^t \Sigma(\xi, x-\xi\Omega,\Omega,\mu)\,d\xi\right). \tag{2.30}$$

It is easily seen from the proof of Theorems 2.1 and 2.2 that the sequence $\{N^{(k)}\}$ given by (2.29) also converges to the unique solution $\tilde{N}$ of (1.2)–(1.4).

In case $\sigma = \Sigma N$, $\sigma_s = \Sigma_s N$, where Σ and $\int_0^{\mu_m}\int_S \Sigma_s(\cdot,\Omega',\mu')\,d\Omega'\,d\mu'$ are bounded continuous functions on E, then the hypothesis (H) is fulfilled with $\rho = \Sigma$, $\rho_s = \Sigma_s$. It follows from Theorem 2.2 and the above remark that we obtain the following conclusions for the linear problem (1.1), (1.3), (1.4):

Theorem 2.3: Let $\Sigma(t,x,\Omega,\mu)$ and $\int_0^{\mu_m}\int_S \Sigma_s(t,x,\Omega,\Omega',\mu,\mu')\,d\Omega'\,d\mu'$ be bounded continuous on E. Then the sequence $\{N^{(k)}\}$ given either by (2.25) or by (2.29) converges to the same unique solution $\tilde{N}$ of the linear problem (1.1), (1.3), (1.4), where f_0, g are defined, respectively, by (2.4) and (2.28) with $\sigma = \Sigma N$, $\sigma_s = \Sigma_s N$.

In Theorems 2.1 and 2.2 the functions σ, σ_s are assumed to satisfy the global Lipschitz conditions (2.2), (2.3). When σ, σ_s satisfy only local Lipschitz condition, that is, for some constant $M>0$,

$$\begin{aligned} &|\sigma(t,x,\Omega,\mu,\eta_1) - \sigma(t,x,\Omega,\mu,\eta_2)| \le \rho_M(t,x,\Omega,\mu)\,|\eta_1-\eta_2| \\ &|\sigma_s(t,x,\Omega,\Omega',\mu,\mu',\eta_1) - \sigma_s(t,x,\Omega,\Omega',\mu,\mu',\eta_2)| \\ &\quad \le \rho_M^*(t,x,\Omega,\Omega',\mu,\mu')\,|\eta_1-\eta_2|, \\ &\eta_1,\eta_2 \in [-M,M], \end{aligned} \tag{2.31}$$

where ρ_M, ρ_M^* are nonnegative functions satisfying

$$\bar{\rho}_M = \sup_{(t,x,\Omega,\mu)\in E}\left(\rho_M(t,x,\Omega,\mu) + \int_0^{\mu_m}\int_S \rho_M^*(t,x,\Omega,\Omega',\mu,\mu')\,d\Omega'\,d\mu'\right) < \infty, \tag{2.32}$$

then it is possible to show the existence of a local solution $\tilde{N}$ in the sense that, for some $t_1>0$, $\tilde{N}(t,x.\Omega,\mu)$ satisfies the system (1.2)–(1.4) for $t \in [0,t_1]$, $(x,\Omega,\mu) \in Q \equiv D\times S\times[0,\mu_m]$. In fact, we have the following:

Theorem 2.4: Assume that $\sigma = \sigma_s = 0$ when $N=0$ and that (2.31) holds for some $M>0$. Then for $|\phi(x,\Omega,\mu)| < M$, the problem (1.2)–(1.4) has a unique local solution $\tilde{N}(t,x,\Omega,\mu)$ on $[0,t_1]\times Q$ for some $t_1>0$, where $[0,t_1]$ is the largest interval in which $|\tilde{N}|$ is bounded by M. If (2.31) holds for every finite M, where ρ_M, ρ_M^* may depend on M, then $\tilde{N}$ can be continued for as long as it remains bounded on Q.

Proof: Define modifications for σ, σ_s by

$$\tilde{\sigma}(t,x,\Omega,\mu,\eta) = \begin{cases} \sigma(t,x,\Omega,\mu,M) & \text{if } \eta > M, \\ \sigma(t,x,\Omega,\mu,\eta) & \text{if } |\eta| \le M, \\ \sigma(t,x,\Omega,\mu,-M) & \text{if } \eta < -M, \end{cases} \tag{2.33}$$

and a similar expression for $\tilde{\sigma}_s(t,x,\Omega,\Omega',\mu,\mu',\eta)$. Then

$\tilde{\sigma}$, $\tilde{\sigma}_s$ are continuous in η and satisfy the respective conditions (2.2), (2.3) for all $\eta_1, \eta_2 \in (-\infty, \infty)$ (with $\rho = \rho_M$, $\rho_s = \rho_M^*$). Hence by Theorem 2.2 the "modified problem" (1.2)–(1.4), where σ, σ_s are replaced by $\tilde{\sigma}, \tilde{\sigma}_s$, has a unique solution $\tilde{N}(t, x, \Omega, \mu)$ which can be constructed from (2.25). Since $|\tilde{N}(0, x, \Omega, \mu)| = |\phi(x, \Omega, \mu)| < M$, there exists $t_1 > 0$ such that $|\tilde{N}(t, x, \Omega, \mu)| \leq M$ on $[0, t_1] \times Q$. But $\tilde{\sigma}, \tilde{\sigma}_s$ coincide, respectively, with σ, σ_s, when $|\tilde{N}(t, x, \Omega, \mu)| \leq M$, we see that $\tilde{N}$ is the unique solution of the original problem (1.2)–(1.4) for at least $(t, x, \Omega, \mu) \in [0, t_1] \times Q$, and it remains so whenever $|\tilde{N}| \leq M$. In case (2.31) holds for every finite M, we may consider $\tilde{N}(t_1, x, \Omega, \mu)$ as the new initial function ϕ in (1.3) and choose $M_1 > M$ in the definition of $\tilde{\sigma}, \tilde{\sigma}_s$ in (2.33). Then the same argument leads to a unique local solution of (1.2)–(1.4) for $t \in [t_1, t_2]$, $(x, \Omega, \mu) \in Q$, where $t_2 > t_1$. By continuing this process and using the uniqueness property of the solution, we can continue the solution for as long as it remains bounded on Q. This proves the theorem.

Theorem 2.4 states that under the condition (2.31) for every finite M the local solution either exists on the whole domain E or is unbounded at some finite $t \in (0, T]$ and some point $(x, \Omega, \mu) \in Q$. This result is in analogy to the classical existence theorem for Cauchy problems of ordinary differential equations. Furthermore, the proof of Theorem 2.4 shows that the local solution $\tilde{N}$ can be constructed through the recursion formula (2.25). However, in the process of construction the functions σ, σ_s in the definition of f_0 should be replaced by $\tilde{\sigma}, \tilde{\sigma}_s$ since it is not known whether the sequence $\{N^{(k)}\}$ remains convergent if the original functions σ, σ_s are used. It is to be noted that the conditions in (2.31) are satisfied by any polynomial function in N with bounded coefficients. For example, if

$$\sigma(t, x, \Omega, \mu, N) = a_1(t, x, \Omega, \mu)N + a_2(t, x, \Omega, \mu)N^2, \tag{2.34}$$

where a_1, a_2 are bounded continuous functions on E then the first condition in (2.31) is satisfied with $\rho_M = a_1 + 2Ma_2$. The second condition in (2.31) is also satisfied if σ_s has the same form as in (2.34).

3. POSITIVE SOLUTION

In this section we show that the solution $\tilde{N}$ of the problem (1.2)–(1.4) is positive or at least nonnegative whenever the initial function ϕ and the source q are nonnegative. It is easily shown by induction that if σ, σ_s are given by the linear function $\sigma = \Sigma N$, $\sigma_s = \Sigma_s N$, where Σ, Σ_s are nonnegative, and if the initial iteration $N^{(0)}$ is nonnegative, then each approximation $N^{(k)}$ given by (2.29) is also nonnegative and thus by Remark 2.1(b) the sequence $\{N^{(k)}\}$ converges to a nonnegative solution $\tilde{N}$. In fact, if we take $N^{(0)} = 0$, then $\{N^{(k)}\}$ converges monotonically to $\tilde{N}$. In view of the uniqueness theorem the solution of the linear problem (1.1), (1.3), (1.4) is therefore nonnegative as is to be expected from physical point of view. The same conclusion holds if $\sigma = \Sigma N$ and σ_s is nonlinear in N but is a nondecreasing function of N. This is due to the fact that the sequence $\{N^{(k)}\}$ given by (2.29) is nonnegative for each k since the function g defined in (2.28) is nondecreasing in N and is nonnegative when $N \geq 0$. The monotone property of this sequence also follows from the nonincreasing property of g when $N^{(0)} = 0$. However, in the general case, where both σ, σ_s are nonlinear in N, the sequence of approximations is given by (2.25) in which the function f_0 is not necessarily nonnegative, and thus this sequence may not be nonnegative for every k. To overcome this difficulty, we transform the problem (1.2)–(1.4) by letting $u = \exp[-(\alpha - \beta)t]N$, where $\alpha > 2\beta \geq 2\bar{\rho}$. Then (1.2) is transformed to the form (with $v = 1$)

$$\frac{\partial u}{\partial t} + \Omega \cdot \nabla u + \alpha u = f(t, x, \Omega, \mu, u) + \beta u, \tag{3.1}$$

where f is given by (2.8) with $\lambda = (\alpha - \beta)$. As usual, the boundary and initial conditions (1.3), (1.4) are reduced to (2.6), (2.7), respectively. In view of (3.1) which is of the same form as (2.5) except with f replaced by

$$f_\beta(t, x, \Omega, \mu, u) = f(t, x, \Omega, \mu, u) + \beta u \tag{3.2}$$

we can obtain a sequence $\{u^{(k)}\}$ from the recursion formula

$$\begin{aligned} u^{(k)}(t, x, \Omega, \mu) = {} & \exp(-\alpha t)\,\phi(x - \Omega t, \Omega, \mu) + \int_0^t \exp[-\alpha(t - \tau)] \\ & \times f_\beta(\tau, x - \Omega(t - \tau), \Omega, \mu, \\ & u^{(k-1)}(\tau, x - \Omega(t - \tau), \Omega, \mu))\, d\tau \\ & k = 1, 2, \cdots. \end{aligned} \tag{3.3}$$

Since by the hypothesis (H) the function

$$(f_\beta(u))(t, x, \Omega, \mu) = f_\beta(t, x, \Omega, \mu, u(t, x, \Omega, \mu)) \tag{3.4}$$

satisfies the condition [see (2.21)]

$$\|f_\beta(w_1) - f_\beta(w_2)\| \leq (\beta + \bar{\rho})\|w_1 - w_2\| \quad [w_1, w_2 \in C_0(E)] \tag{3.5}$$

and since $\alpha > \beta + \bar{\rho}$, we see from the proof of Theorem 2.1 that the sequence $\{u^{(k)}\}$ given by (3.3) converges to a unique solution $\tilde{u}$ of the problem (3.1), (2.6), (2.7) and this solution must be the same as that given in Theorem 2.1 (with $\lambda = \alpha - \beta$). The implication of this new sequence is that it leads to a monotone nondecreasing sequence as is shown in the following:

Theorem 3.1: Let (H) be satisfied and let ϕ, q be nonnegative on E. If $\sigma_s \geq 0$ when $N \geq 0$, then for any $u^{(0)} \geq 0$ the sequence $\{u^{(k)}\}$ given by (3.3) converges in $C_0(E)$ to a unique nonnegative solution $\tilde{u}$ of the problem (3.1), (2.6), (2.7). If, in addition, σ_s is nondecreasing in N and $u^{(0)} \equiv 0$, then $u^{(k-1)} \leq u^{(k)}$ for every k and $\{u^{(k)}\}$ converges monotonically to $\tilde{u}$.

Proof: For the first part of the theorem it suffices to show that $u^{(k)} \geq 0$ on E for every k since the convergence of $\{u^{(k)}\}$ follows from (3.5) and the contraction mapping theorem. By the condition (2.2) with $\eta_1 = u^{(k-1)}$, $\eta_2 = 0$ and by the hypothesis on σ_s we have

$$\begin{aligned} & |\sigma(t, x, \Omega, \mu, u^{(k-1)})| \leq \rho(t, x, \Omega, \mu)\,|u^{(k-1)}| \leq \bar{\rho} u^{(k-1)}, \\ & \sigma_s(t, x, \Omega, \Omega', \mu, \mu', u^{(k-1)}) \geq 0 \end{aligned} \tag{3.6}$$

when $u^{(k-1)} \geq 0$. The above inequalities together with the nonnegative property of q imply that, for $u^{(k-1)} \geq 0$ on E,

$$f(t, x, \Omega, \mu, u^{(k-1)}) + \beta u^{(k-1)} \geq (\beta - \bar{\rho})u^{(k-1)} \geq 0 \tag{3.7}$$

[see (2.4), (2.8)]. Hence by (3.3) and the nonnegative

property of ϕ and $u^{(0)}$ we have $u^{(1)} \geq 0$ on E. The proof of $u^{(k)} \geq 0$ on E for every k follows from (3.3), (3.7) by induction. To show the monotone property of $\{u^{(k)}\}$ when σ_s is nondecreasing in N, we observe from the choice of $u^{(0)} \equiv 0$ that $u^{(1)} \geq u^{(0)}$. Assume, by induction, that $u^{(k)} \geq u^{(k-1)}$ on E. Then since

$$\sigma_s(t, x, \Omega, \Omega', \mu, \mu', \exp(\lambda t) u^{(k)}(t, x, \Omega', \mu')) \geq \sigma_s(t, x, \Omega, \Omega', \mu, \mu' \exp(\lambda t) u^{(k-1)}(t, x, \Omega', \mu')),$$

where $\lambda = (\alpha - \beta)$, we obtain from (2.2) that

$$f_\beta(t, x, \Omega, \mu, u^{(k)}) - f_\beta(t, x, \Omega, \mu, u^{(k-1)}) \geq (\beta - \bar{\rho})(u^{(k)} - u^{(k-1)}) \geq 0. \tag{3.8}$$

It follows from (3.3) that $u^{(k+1)} - u^{(k)} \geq 0$. This completes the proof of the theorem.

By letting $N^{(k)} = \exp[(\alpha - \beta)t] u^{(k)}$, $N = \exp[(\alpha - \beta)t] u$, and

$$f_\beta^*(t, x, \Omega, \mu, N) = f_0(t, x, \Omega, \mu, N) + \beta N, \tag{3.9}$$

we obtain the following conclusion:

Theorem 3.2: Let the conditions in Theorem 3.1 be satisfied and let $N^{(0)} \geq 0$. Then the sequence $\{N^{(k)}\}$ given by

$$N^{(k)}(t, x, \Omega, \mu) = \exp(-\beta t)\, \phi(x - \Omega t, \Omega, \mu) + \int_0^t \exp[-\beta(t - \tau)] \times f_\beta^*(\tau, x - \Omega(t - \tau), \Omega, \mu, N^{(k-1)}(\tau, x - \Omega(t - \tau), \Omega, \mu))\, d\tau, \quad k = 1, 2, \cdots, \tag{3.10}$$

converges to a unique nonnegative solution $\tilde{N}$ of the problem (1.2)–(1.4). If, in addition, σ_s is nondecreasing in N and $N^{(0)} \equiv 0$, then $N^{(k)} \geq N^{(k-1)}$ on E for every k and $\{N^{(k)}\}$ converges monotonically to $\tilde{N}$.

Proof: Since the nonnegativity and monotonicity as well as the convergence of $\{u^{(k)}\}$ imply the same for $\{N^{(k)}\}$ and since (3.3) is reduced to (3.10) when $u^{(k)}$ is replaced by $\exp[-(\alpha - \beta)t] N^{(k)}$, the conclusions in the theorem follow immediately from Theorem 3.1.

Remark 3.1: The assumption of the nondecreasing property of σ_s in Theorems 3.1 and 3.2 is physically realistic since σ_s represents the neutrons production due to scattering and fission. It is obvious that more neutrons in the medium will cause more scattering and fission.

[1] K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass. 1967).
[2] K. M. Case, SIAM—AMS Proc. 1, 17 (1969).
[3] R. Bellman, SIAM—AMS Proc. 1, 95 (1969).
[4] R. Bellman, R. Kalaba, and G. M. Wing, J. Math. Mech. 7, 741 (1958).
[5] J. Lehner and G. M. Wing, Duke Math. J. 23, 125 (1956).
[6] G. M. Wing, *An Introduction to Transport Theory* (Wiley, New York, 1962).
[7] B. Davison, *Neutron Transport Theory* (Oxford U. P., London, 1957).
[8] H. Hejtmanek, J. Math. Phys. 11, 995 (1970).
[9] E. Inönü and P. F. Zweifel, Eds., *Developments in Transport Theory* (Academic, New York, 1967).
[10] C. V. Pao, Arch. Rat. Mech. Anal. 50, 290 (1973).
[11] C. V. Pao, Trans. Am. Math. Soc. 194, 167 (1974).
[12] A. Leonard and J. H. Ferziger, Nucl. Sci. Eng. 26, 170 (1966).
[13] K. W. Reed, J. Anal. Math. 17, 347 (1966).
[14] A. Suhadolc, J. Math. Anal. Appl. 35, 1 (1971).
[15] I. Vidav, J. Math. Anal. Appl. 30, 264 (1970).
[16] J. Mika, J. Math. Phys. 15, 892 (1974).

The Hamiltonians of the Schrödinger algebra

G. Burdet and M. Perrin

Centre National de la Recherche Scientifique, 13274 Marseille Cedex 2, France*
(Received 20 January 1975)

We give a complete characterization of the one-dimensional nonconjugate subalgebras of the extended Schrödinger algebra (and of the Schrödinger algebra). This yields a classification into conjugacy classes of all the Hamiltonians admitting the extended Schrödinger algebra as time dependent invariance algebra in the Schrödinger picture. Moreover, we determine the action of the inner automorphism induced by the general element of the extended Schrödinger algebra from which the explicit time dependence of the generators is deduced, whatever Hamiltonian is considered.

INTRODUCTION

This paper is a continuation of a series[1] devoted to the study of possibilities offered by the introduction of explicit time-dependent operators in the Schrödinger picture of the Galilean quantum mechanics. Indeed, given H the Hamiltonian of a conservative system, the Schrödinger equation $H\psi = i\,\partial/\partial t\,\psi$ transforms covariantly under the time-dependent transformation generated by the operator $S(t)$ if the Hamiltonian transforms according to $H \to \hat{H} = S(t)HS^{-1}(t) + i[\partial S(t)/\partial t]S^{-1}(t)$ whereas ψ cotransforms following $\psi \to \hat{\psi} = S(t)\psi$.

Then it is possible to extract from the set of the $S(t)$ operators the ones such that $[H, S(t)] = i\,\partial S(t)/\partial t$; these operators correspond to transformations which keep invariant the Hamiltonian and generate a Lie algebra $\mathcal{S}_t$, not necessarily a finite-dimensional one. Moreover, such operators do not depend explicitly on time in the Heisenberg picture and take their values for $t = 0$; we set

$$S \equiv S_H = \exp(iHt)S(t)\exp(-iHt) = S(0).$$

Consequently, the set $\{S\}$ is a Lie algebra $\mathcal{S}_0$ isomorphic to $\mathcal{S}_t$, which makes clear the meaning of the time t as a parameter in the Galilean mechanics.

Conversely, any Lie algebra of time-independent operators which do not commute with a Hamiltonian in the Heisenberg picture corresponds to a time-dependent invariance algebra in the Schrödinger picture.

It is worth noticing that the generators of such invariance algebras are integrals of motion owing to the vanishing of the time derivatives of their expectation values.

Now let $\mathcal{E}(\Omega)$ be the enveloping algebra of the fundamental observables $\Omega = \{\mathbf{q}(\mu), \mathbf{p}(\mu), m(\mu)\}$ characterizing the individual constituents ($\mu \in [1, n]$) of a quantum mechanical system,[2] from the above properties every Lie algebra which can be mapped via an injective homomorphism into $\mathcal{E}(\Omega)$ can be considered as an invariance algebra for any one of its one-dimensional subalgebras interpreted as a Hamiltonian.

In this paper the case of the one central element extension $\widetilde{\mathcal{S}}_{ch_3}$ of $\mathcal{S}_{ch_3}$, the Schrödinger algebra, is treated, and likewise the Galilean case since $\widetilde{\mathcal{G}} \subset \widetilde{\mathcal{S}_{ch_3}}$.

In the first section a complete characterization of the one-dimensional nonconjugate subalgebras of $\widetilde{\mathcal{S}_{ch_3}}$ (and of $\mathcal{S}_{ch_3}$) is given. Indeed, the eigenstates of two Hamiltonians conjugate by an inner automorphism of $\widetilde{\mathcal{S}_{ch_3}}$ are naturally related, so it is sufficient to classify the Hamiltonians into conjugacy classes. Hence we obtain a reliable classification of all the Hamiltonians possesing the Schrödinger algebra as invariance algebra. It must be noticed that the same kind of classification takes place in the problem of separation of variables in the Schrödinger equation.[3] Indeed it has been shown in Ref. 3 that the possible coordinates which permit the separation are associated to the orbits of the Schrödinger group on the projective space of the Schrödinger algebra.

In the second part the explicit time dependence carried by the generators of the Schrödinger algebra in the Schrödinger picture is exhibited for all the possible Hamiltonians. This is achieved by computing the action of the inner automorphism induced by the general element of $\widetilde{\mathcal{S}}_{ch_3}$ on the dual vector space of $\widetilde{\mathcal{S}}_{ch_3}$ (i.e., we make use of the coadjoint representation).

I. ONE-DIMENSIONAL NONCONJUGATE SUBALGEBRAS OF $\mathcal{S}ch_3$ AND $\widetilde{\mathcal{S}ch_3}$

The extended Schrödinger algebra can be written as

$$\widetilde{\mathcal{S}}_{ch_3} \cong \mathcal{H}_3 \square (\mathcal{S}O_3 \oplus \mathcal{S}O_{2,1}) \tag{1}$$

where $\mathcal{H}_3$ is the Heisenberg algebra (i.e., the Lie algebra of the Weyl group) and $\mathcal{S}O_3$ is the familiar rotational algebra. We use a natural basis which comes from the above decomposition, a physical realization of which, in terms of the canonical variables, and the nonzero commutation relations are given by

$$\mathcal{H}_3: \quad \left[\begin{array}{ll} K_j = \sum_\mu m(\mu) q_j(\mu) & \\ P_j = \sum_\mu p_j(\mu) & [K_j, P_k] = i\delta_{jk} M \\ M = \sum_\mu m(\mu) & \end{array}\right] \tag{2a}$$

$$\mathcal{S}O_3: \quad \left[J_j = \sum_\mu \epsilon_{jkl} q_k(\mu) p_l(\mu) \qquad [J_j, J_k] = i\epsilon_{jkl} J_l\right] \tag{2b}$$

$$\mathcal{S}O_{2,1}: \quad \left[\begin{array}{ll} T = \frac{1}{2}\sum_\mu \frac{p^2(\mu)}{m(\mu)} & [T, C] = iD \\ C = \frac{1}{2}\sum_\mu m(\mu) q^2(\mu) & [D, T] = -2iT \\ D = -\frac{1}{2}\sum_\mu [\mathbf{q}(\mu), \mathbf{p}(\mu)]_+ & [D, C] = 2iC \end{array}\right]. \tag{2c}$$

The action of the semisimple part $\mathcal{S}O_3 \oplus \mathcal{S}O_{2,1}$ on the solvable ideal $\mathcal{H}_3$ leads to the following nonzero commutation relation

$$[J_j, K_k] = i\epsilon_{jkl} K_l, \quad [T, K_j] = -iP_j, \quad [D, K_j] = iK_j,$$
$$[J_j, P_k] = i\epsilon_{jkl} P_l, \quad [C, P_j] = iK_j, \quad [D, P_j] = -iP_j. \tag{3}$$

The extended Galilei algebra

$$\tilde{\mathcal{G}}_3 \cong \mathcal{H}_3 \square (\mathcal{S}O_3 \oplus \mathfrak{A}_1(T)) \tag{4}$$

where $\mathfrak{A}_1(T)$ is the noncompact one-dimensional algebra generated by T, clearly appears as a subalgebra of $\widetilde{\mathcal{S}_{ch_3}}$.

Obviously a six-dimensional Abelian algebra $\mathfrak{A}_6 \tilde{=} \mathcal{H}_3/\mathfrak{A}_1(M)$ takes the place of the Heisenberg algebra in the corresponding nonextended algebras $\mathcal{S}_{ch_3}$ and $\mathcal{G}_3$.

The general element of $\widetilde{\mathcal{S}_{ch_3}}$, reflecting the vector space decomposition of the rhs of (1), is denoted

$$X = \lambda \cdot \mathbf{J} + \delta T + \beta C + \gamma D + \rho \cdot \mathbf{P} + \tau \cdot \mathbf{K} + \mu M \tag{5}$$

where $\delta, \beta, \gamma, \mu \in \mathbb{R}$ and $\lambda, \rho, \tau \in \mathbb{R}^3$. Now let us consider the problem of determining the one-dimensional nonconjugate subalgebras of $\widetilde{\mathcal{S}_{ch_3}}$. Owing to the central extension structure it is easy to deduce that two nonconjugate subalgebras of $\mathcal{S}_{ch_3}$ remain nonconjugate subalgebras in $\widetilde{\mathcal{S}_{ch_3}}$. Consequently, we have first to consider the one-dimensional nonconjugate subalgebras of $\mathcal{S}_{ch_3}$. Then the semidirect sum structure makes this task easier. Indeed let $\mathcal{A}$ be the semidirect sum $\mathcal{A} \cong \mathcal{B} \square_\sigma \mathcal{C}$ [defined by the homomorphism $\sigma: \mathcal{C} \to \mathcal{D}(\mathcal{B})$ the derivation algebra of B], then two nonconjugate subalgebras of $\mathcal{C}$ remain nonconjugate as subalgebras of $\mathcal{A}$. In the decomposition of $\widetilde{\mathcal{S}_{ch_3}}$ given by (1) $\mathcal{C}$ is the semisimple part $\mathcal{S}O_3 \oplus \mathcal{S}O_{2,1}$, the one-dimensional nonconjugate subalgebras of which are well known: $\mathcal{S}O_3$ possesses only one-dimensional subalgebra corresponding to the fact that λ^2 its continuous invariant is definite positive; $\mathcal{S}O_{2,1}$ possesses three one-dimensional nonconjugate subalgebras according to whether the invariant $\alpha^2 = \gamma^2 - \beta\delta$ is positive, negative, or null. The coupling between the $\mathcal{S}O_3$ and $\mathcal{S}O_{2,1}$ one-dimensional subalgebras causes to appear families dependent on the parameter $\eta = \frac{|\alpha|}{|\lambda|}$ (i.e., to two distinct values of the parameter correspond two distinct classes).

It then remains to determine how a nontrivial $\mathfrak{A}_6$

TABLE I. One-dimensional nonconjugate subalgebras (ODNCS) of the one central element extended Schrödinger algebra.[a]

	ODNCS of the Schrödinger algebra						
Family or class	Representative elements—part I	Characterization of the various components				Representative elements—part II	Characterization of the components and parameters
		$\mathcal{S}O_3$	$\mathcal{S}O_{2,1}$	$\mathfrak{A}_3(P) \mathfrak{A}_3(K)$	$\mathfrak{A}_1(M)$		
1	J_3	$\lvert\lambda\rvert>0$	×	$\lambda\cdot\rho=\lambda\cdot\tau=0$	$\mu=\mu_0$	$J_3+\xi M$	$\mu\neq\mu_0=\frac{(\lambda,\tau,\rho)}{\lambda^2}$
2	J_3+K_3	$\lvert\lambda\rvert>0$	×	$\rho,\tau\in\mathbb{R}^3$	$\mu\in\mathbb{R}$		
3	$T+C$	×	$\alpha^2<0$	$\rho,\tau\in\mathbb{R}^3$	$\mu=\mu_0$	$T+C+\xi_0 M$	$\mu\neq\mu_0=-\Xi/2\alpha^2$
4	D	×	$\alpha^2>0$	$\rho,\tau\in\mathbb{R}^3$	$\mu=\mu_0$	$D+\xi_0 M$	$\mu\neq\mu_0=-\Xi/2\alpha^2$
5	T	×	$\alpha^2=0$	$\delta^{-1/2}\rho=\beta^{-1/2}\tau=\nu\in\mathbb{R}^3$	$\mu=\mu_0$	$T\pm M$	$\mu\neq\mu_0=\frac{1}{2}\nu^2$
6	$T+K_3$	×	$\alpha^2=0$	$\rho,\tau\in\mathbb{R}^3$	$\mu\in\mathbb{R}$		
7_η	$J_3+\eta(T+C)$	$\lvert\lambda\rvert>0$	$\alpha^2<0$	$\rho,\tau\in\mathbb{R}^3$	$\mu=\mu_0$	$J_3+\eta(T+C)+\xi M$	$\mu\neq\mu_0=\frac{-1}{2(\alpha^2+\lambda^2)^2}[(\alpha^2+3\lambda^2)\Xi+\left(\frac{\lambda^2}{\alpha^2}-1\right)\kappa-4\lambda^2(\lambda,\tau,\rho)]$
8	J_3+T+C	$\lvert\lambda\rvert>0$ $\lvert\lambda\rvert=\lvert\alpha\rvert$	$\alpha^2<0$	$\lambda\wedge\rho=\lambda\wedge\tau=0$	$\mu=\mu_0$	$J_3+T+C+\xi M$	$\mu\neq\mu_0=-\frac{\kappa}{\alpha^2}$
9ζ	$J_3+T+C+\zeta K_2$	$\lvert\lambda\rvert>0$ $\lvert\lambda\rvert=\lvert\alpha\rvert$	$\alpha^2<0$	$\rho,\tau\in\mathbb{R}^3$	$\mu=\mu_0$	$J_3+T+C+\zeta K_2+\xi M$	$\mu\neq\mu_0=\frac{\kappa}{2\lambda^4}+\frac{(\lambda,\tau,\rho)}{2\lambda^2}$, $\zeta^2=\frac{\Xi}{\sqrt{-\alpha^2}}-\frac{\kappa}{\lambda^2\sqrt{-\alpha^2}}+2\frac{(\lambda,\tau,\rho)}{\lvert\lambda\rvert}$
10η	$J_3+\eta D$	$\lvert\lambda\rvert>0$	$\alpha^2>0$	$\rho,\tau\in\mathbb{R}^3$	$\mu=\mu_0$	$J_3+\eta D+\xi M$	$\mu\neq\mu_0=\frac{-1}{2(\alpha^2+\lambda^2)}[\Xi+\frac{\kappa}{\alpha^2}-2(\lambda,\tau,\rho)]$
11	J_3+T	$\lvert\lambda\rvert>0$	$\alpha^2=0$	$\lambda\cdot\sigma=0$	$\mu=\mu_0$	$J_3+T+\xi M$	$\mu\neq\mu_0=\frac{(\lambda,\tau,\rho)}{\lambda^2}+\frac{1}{2\lvert\lambda\rvert}(\nu_\lambda^2-\sigma^2)$.
12ξ	$J_3+T+\xi K_3$	$\lvert\lambda\rvert>0$	$\alpha^2=0$	$\rho,\tau\in\mathbb{R}^3$	$\mu\in\mathbb{R}$		$\xi=\lambda\cdot\sigma/\lvert\lambda\rvert^{5/2}$
13	K_3	×	×	$\rho\wedge\tau=0$	$\mu\in\mathbb{R}$		
14	K_3+P_2	×	×	$\rho,\tau\in\mathbb{R}^3$	$\mu\in\mathbb{R}$		
15		×	×	$\rho=\tau=0$		M	$\mu=1$

[a] We set: $\Xi=\beta\rho^2+\delta\tau^2+2\gamma(\rho\cdot\tau)$, $\chi=\beta(\rho\cdot\lambda)^2+\delta(\tau\cdot\lambda)^2+2\gamma(\rho\cdot\lambda)(\tau\cdot\lambda)$, $\xi=(\mu-\mu_0)/\lvert\lambda\rvert$, $\xi_0=(\mu-\mu_0)/\lvert\alpha\rvert$, $\sigma=\beta^{1/2}\rho+\delta^{1/2}\tau$, $\nu_\lambda=(\nu\cdot\lambda)/\lvert\lambda\rvert$.

component can be added; this yields some constraints on the parameters ρ and τ in the full characterization of the one-dimensional nonconjugate subalgebras of $\mathcal{S}_{ch_3}$ and causes the appearance of new families indexed by ζ. The results are summarized in Table I (part I) where a representative element of each conjugacy class is given. Note that it must be distinguished between the case $\alpha^2=0$ with $\beta=\gamma=\delta=0$ (denoted by crossed boxes in Table I) and the nontrivial $\alpha^2=0$. Obviously such a distinction is meaningless for λ^2.

Next the one-dimensional nonconjugate subalgebras of $\widetilde{\mathcal{S}_{ch_3}}$ can be deduced by computing how the mass generator adds to each subalgebra of $\mathcal{S}_{ch_3}$, causing the appearance of families indexed by ξ, without omitting M which is its own conjugacy class. The results are gathered in Table I (parts I and II).

It is interesting to compare the one-dimensional nonconjugate subalgebras of the Schrödinger algebra to the Galilean ones.[4] The following classes and families 1, 2, 5, 6, 11, 12_ζ, 13, 14 appear in the Galilei algebra also. One class represented by P_3 and one family $J_3+\zeta P_3$ belonging to the one-dimensional nonconjugate subalgebras of $\mathcal{G}$ are conjugate to K_3 and J_3+K_3, respectively, hence do not appear in Table I. Furthermore, by the existence of a dilation, generated by D, the following four families disappeared: $J_3+\zeta K_3$, $T+\zeta K_3$, $J_3+\eta T$, $K_3+\rho_2 P_2$, the four classes 2, 6, 11, and 14 taking their place, respectively.

In Table II the centralizers of the one-dimensional subalgebras of $\widetilde{\mathcal{S}_{ch_3}}$ (and $\mathcal{S}_{ch_3}$) are given. In quantum mechanics they are interpreted as symmetry algebras, the generators of which, commuting with the respective Hamiltonian, do not take any time dependence.

What about the one-dimensional subalgebras of $\widetilde{\mathcal{S}_{ch_3}}$ as Hamiltonians? By construction all the subalgebras belonging to the part II in Table I possess a nonzero component on M, so they cannot be considered as Hamiltonians in Galilean mechanics since a rest mass term is not contained in the energy. Consequently, the candidates as Hamiltonians stand in the part I of Table I only. Let us note that this part contains the common $\mathcal{S}_{ch_3}$ and $\mathcal{S}_{ch_3}$ subalgebras, recalling the position expounded in Ref. 3.

One finds again[1b] three familiar Hamiltonians as representative elements of the classes 3, 4, and 5, which are, respectively:

$T+\omega^2 C$: the Hamiltonian of a system in an external isotropic harmonic field;

$T-\omega^2 C$: the Hamiltonian of a system in an external "antiharmonic" isotropic field;

T: The Galilean Hamiltonian of a free (isolated) system.

Other representatives of the classes 3 and 4 are $T \pm \omega^2 C+(e/m)\mathbf{E}\cdot\mathbf{K}$ which are the Hamiltonians of a system of identical charged (e) particles in an uniform electric field $\mathbf{E}$ besides the (anti)harmonic one.

The Hamiltonian of a system with a linear potential $T+\phi\mathbf{K}$ [free fall or dipolar coupling of a system in an uniform electric field (Stark effect)] is a representative of the class 6.

In the class 1 one finds J_3 which governs a system having a magnetic momentum in a static magnetic field to which a transversal uniform electric field can be added. The same physical system furnishes representative Hamiltonians of the class 2, in which an uniform electric field cannot be excluded.

Let us now consider the Hamiltonian of an identical charged particle system of spin zero in an uniform magnetic field $\mathbf{B}$ directed along the 3-axis ($B\equiv B_3$) which can be written

$$\frac{1}{2m}\sum_\mu p^2(\mu)+\frac{e}{c}B(\mathbf{q}(\mu)\wedge\mathbf{p}(\mu))_3+\frac{e^2}{4c^2}B^2(q_1^2(\mu)+q_2^2(\mu))$$

Such a Hamiltonian cannot be expressed in terms of generators of $\widetilde{\mathcal{S}_{ch_3}}$, but this impossibility can be overcome by adding a convenient anisotropic external harmonic field such that

$$\omega_1^2=\omega_2^2=\omega_T^2 \quad \text{and} \quad \frac{e^2}{4m^2c^2}B^2+\omega_T^2=\omega_3^2=\omega^2;$$

then we obtain the Hamiltonians

$$H=T+\frac{e}{2mc}BJ_3+\omega^2 C$$

which are representatives of the family 7_η with $\eta=2mc|\omega|/eB$.

Moreover, we get a representative of the class 8 if $\omega_T=0$. Again a uniform electric field can be added in the above Hamiltonians as long as the constraints on τ given in Table I are fulfilled. As in class 2, a uniform electric field cannot be eliminated in the representatives of the family 9ζ.

II. THE SCHRÖDINGER ALGEBRA IN THE SCHRÖDINGER PICTURE

To every Hamiltonian H admitting $\widetilde{\mathcal{S}_{ch_3}}$ as an invariance algebra corresponds an explicit time dependence of the generators determined by

$$S(t)=\exp(-iHt)S(O)\exp(iHt). \tag{6}$$

(Obviously, as already mentioned, the stabilizer elements of H do not take a time dependence.)

By taking as a Hamiltonian X the general element of $\widetilde{\mathcal{S}_{ch_3}}$ defined in (5), we obtain the general explicit time-dependent form of the generators of $\widetilde{\mathcal{S}_{ch_3}}$ in the Schrödinger picture at once.

We use the technics described in Ref. 1c. Let us introduce the coadjoint representation of X

$$\mathrm{coad}X=i\Phi \tag{7}$$

where Φ is a 13×13 matrix which is going to act on the dual vector space of the Lie algebra. Then we have just to construct the matrix $\exp(\Phi t)$ since the explicit time dependence is given by

$$\begin{bmatrix} \mathbf{P}(t) \\ \mathbf{K}(t) \\ M \\ \mathbf{J}(t) \\ T(t) \\ C(t) \\ D(t) \end{bmatrix} = e^{\Phi t} \begin{bmatrix} \mathbf{P} \\ \mathbf{K} \\ M \\ \mathbf{J} \\ T \\ C \\ D \end{bmatrix}. \tag{8}$$

This is achieved by using the standard matrix exponentiation methods.[5] As a consequence of the Cayley–Hamilton theorem, $\exp(\Phi t)$ is expressed in terms of the twelve first powers of Φ and can be written, by using the Lagrange–Sylvester interpolation polynomial, as follows:

$$\exp(\Phi t) = \left[1 + \Phi t + \Phi^2 \left(t^2 + 2\sum_j \frac{1}{\sigma_j^2}\right)\right] \prod_j \frac{\sigma_j^2 \mathbb{I} - \Phi^2}{\sigma_j^2} + \sum_j \frac{\Phi^3}{2\sigma_j^4} \left(\prod_{k\neq j} \frac{\sigma_k^2 \mathbb{I} - \Phi^2}{\sigma_j^2 - \sigma_k^2}\right) [e^{\sigma_j t}(\sigma_j \mathbb{I} + \Phi) + e^{-\sigma_j t}(\sigma_j \mathbb{I} - \Phi)] \tag{9}$$

where the σ_j's ($j = 1, 3, 5, 7, 9$) are the following nonzero roots of the characteristic polynomial of Φ:

$$\sigma_1 = \alpha, \quad \sigma_3 = i|\lambda|, \quad \sigma_5 = \alpha + i|\lambda|, \quad \sigma_7 = \alpha - i|\lambda|, \quad \sigma_9 = 2\alpha.$$

(We recall that λ^2 and α^2 are computed from the $\mathcal{S}O_3$ and $\mathcal{S}O_{2,1}$ components of X, respectively.)

Although the final result is analytic when the powers of Φ are explicitly computed, we cannot apply the above expression (9) when several nonzero characteristic roots become equal or when some of them become zero. This feature causes the appearance of numerous particular cases but we shall not give all of them here.

We must mention the case $\lambda^2 = \alpha^2 = 0$ for which the time dependence becomes a polynomial one

$$e^{\Phi t} = \mathbb{I} + \sum_{n=1}^{12} \frac{t^n}{n!} \Phi^n. \tag{10}$$

Another interesting case appears when H belongs to the semisimple part $\mathcal{S}O_3 \oplus \mathcal{S}O_{2,1}$ only, i.e., when we have $H = H_{|\mathcal{S}O_3} + H_{|\mathcal{S}O_{2,1}}$. Owing to the direct sum structure, the time dependence taken by the $\mathcal{S}O_3$ generators is governed by $H_{|\mathcal{S}O_3}$ only; likewise the $\mathcal{S}O_{2,1}$ generators take a time dependence linked to $H_{|\mathcal{S}O_{2,1}}$. Finally, to obtain the time dependence of the Heisenberg generators we just have to superpose the commuting actions of $H_{|\mathcal{S}O_3}$ and $H_{|\mathcal{S}O_{2,1}}$.

Under the part $H_{|\mathcal{S}O_3} = \boldsymbol{\lambda} \cdot \mathbf{J}$ the components of any vector $\mathbf{V}$ under $\mathcal{S}O_3$ become

$$\begin{aligned} V_j(t) &= \exp(iH_{|\mathcal{S}O_3} \cdot t)\, V_j \exp(-iH_{|\mathcal{S}O_{3,1}} \cdot t) \quad (j = 1, 2, 3) \\ &= \left[1 + (\lambda_j^2 - \lambda^2)\left(t^2 - \frac{\cos|\lambda| t}{\lambda^2}\right)\right] V_j \\ &\quad + \sum_{k\neq j} \left[\lambda_j \lambda_k \left(t^2 - \frac{\cos|\lambda| t}{\lambda^2}\right) + \epsilon_{jkl} \frac{\lambda_l}{|\lambda|} \sin|\lambda| t\right] . \end{aligned} \tag{11}$$

In particular, $\mathbf{K}$, $\mathbf{P}$ and $\mathbf{J}$ which are vectors under $\mathcal{S}O_3$ transform according to (11).

Under the part $H_{|\mathcal{S}O_{2,1}} = \delta T + \beta C + \gamma D$, we get

$$\begin{aligned} P_j(t) &= \left(-\frac{\gamma}{\alpha} \operatorname{sh}\alpha t + \operatorname{ch}\alpha t\right) P_j + \frac{\beta}{\alpha} \operatorname{sh}\alpha t K_j, \\ K_j(t) &= \left(\frac{\gamma}{\alpha} \operatorname{sh}\alpha t + \operatorname{ch}\alpha t\right) K_j - \frac{\delta}{\alpha} \operatorname{sh}\alpha t P_j, \end{aligned} \tag{12a}$$

TABLE II. Stabilizers of the one-dimensional subalgebras.[a]

Class or Family	In the Schrödinger algebra	In the extended Schrödinger algebra
1	$((\mathfrak{A}_1(K_3) \oplus \mathfrak{A}_1(P_3)) \square \mathcal{S}O_{2,1}) \oplus \mathcal{S}O_2(J_3)$	$(\mathcal{H}_1(3) \square \mathcal{S}O_{2,1}) \oplus \mathcal{S}O_2(J_3)$
2	$((\mathfrak{A}_1(K_3) \oplus \mathfrak{A}_1(P_3)) \square \mathfrak{A}_1(C)) \oplus \mathcal{S}O_2(J_3)$	$\mathbb{R}(M) \oplus \mathfrak{A}_1(K_3) \oplus \mathfrak{A}_1(C) \oplus \mathcal{S}O_2(J_3)$
3	$\mathcal{S}O_2(T+C) \oplus \mathcal{S}O_3$	$\mathbb{R}(M) \oplus \mathcal{S}O_2(T+C) \oplus \mathcal{S}O_3$
4	$\mathbb{R}(D) \oplus \mathcal{S}O_3$	$\mathbb{R}(M) \oplus \mathbb{R}(D) \oplus \mathcal{S}O_3$
5	$(\mathfrak{A}_3(P) \square \mathcal{S}O_3) \oplus \mathfrak{A}_1(T)$	$\mathbb{R}(M) \oplus (\mathfrak{A}_3(P) \square \mathcal{S}O_3) \oplus \mathfrak{A}_1(T)$
6	$(\mathfrak{A}_3(P) \square \mathcal{S}O_2(J_3)) \oplus \mathfrak{A}_1(T+K_3)$	$\mathbb{R}(M) \oplus (\mathfrak{A}_2(P_1, P_2) \square \mathcal{S}O_2(J_3)) \oplus \mathfrak{A}_1(T+K_3)$
7η	$\mathcal{S}O_2(T+C) \oplus \mathcal{S}O_2(J_3)$	$\mathbb{R}(M) \oplus \mathcal{S}O_2(T+C) \oplus \mathcal{S}O_2(J_3)$
8	$\mathfrak{A}_2(*) \square (\mathcal{S}O_2(T+C) \oplus \mathcal{S}O_2(J_3))$	$\mathcal{H}_1(*) \square (\mathcal{S}O_2(T+C) \oplus \mathcal{S}O_2(J_3))$
9ξ	$\mathfrak{A}_1(J_3 + T + C + \xi K_2)$	$\mathbb{R}(M) \oplus \mathfrak{A}_1(J_3 + T + C + \xi K_2)$
10η	$\mathcal{S}O_2(J_3) \oplus \mathbb{R}(D)$	$\mathbb{R}(M) \oplus \mathcal{S}O_2(J_3) \oplus \mathbb{R}(D)$
11	$\mathfrak{A}_1(P_3) \oplus \mathcal{S}O_2(J_3) \oplus \mathfrak{A}_1(T)$	$\mathbb{R}(M) \oplus \mathcal{S}O_2(J_3) \oplus \mathfrak{A}_1(P_3) \oplus \mathfrak{A}_1(T)$
12ξ	$\mathfrak{A}_1(P_3) \oplus \mathfrak{A}_1(K_3 + \frac{1}{\xi} T) \oplus \mathcal{S}O_2(J_3)$	$\mathbb{R}(M) \oplus \mathfrak{A}_1(K_3 + \frac{1}{\xi} T) \oplus \mathcal{S}O_2(J_3)$
13	$(\mathfrak{A}_3 \oplus \mathfrak{A}_3) \square (\mathcal{S}O_2(J_3) \oplus \mathfrak{A}_1(C))$	$(\mathcal{H}_2(1,2) \oplus \mathfrak{A}_1(3)) \square (\mathcal{S}O_2(J_3) \oplus \mathfrak{A}_1(C))$
14	$\mathfrak{A}_3 \oplus \mathfrak{A}_3$	$\mathcal{H}_1(1) \oplus \mathfrak{A}_1(P_2) \oplus \mathfrak{A}_1(K_3)$
15		$\widetilde{\mathcal{S}\mathrm{ch}_3}$

[a] The generators corresponding to the particular representative elements given in Table I are written between brackets when they are questionable. In Class 8 we have set $\mathfrak{A}_2(*) \cong \{K_1 - P_2,\ K_2 + P_1\}$ and $\mathcal{H}_1(*) \cong \{K_1 - P_2, K_2 + P_1, M\}$.

$$T(t)=\left[1+\left(1+\frac{\beta\delta}{2\alpha^2}\right)(\cosh 2\alpha t-1)-\frac{\gamma}{\alpha}\sinh 2\alpha t\right]T$$
$$+\frac{\beta^2}{2\alpha^2}(\cosh 2\alpha t-1)C+\left(\frac{\beta\gamma}{2\alpha^2}(\cosh 2\alpha t-1)\right.$$
$$\left.-\frac{\beta}{2\alpha}\sinh 2\alpha t\right)D,$$
$$C(t)=\left[1+\left(1+\frac{\beta\delta}{2\alpha^2}\right)(\cosh 2\alpha t-1)+\frac{\gamma}{\alpha}\sinh 2\alpha t\right]C$$
$$+\frac{\delta^2}{2\alpha^2}(\cosh 2\alpha t-1)T+\left(\frac{\gamma\delta}{2\alpha^2}(\cosh 2\alpha t-1)\right.$$
$$\left.+\frac{\delta}{2\alpha}\sinh 2\alpha t\right)D, \tag{12b}$$
$$D(t)=\left(\frac{\gamma^2}{\alpha^2}-\frac{\beta\delta}{\alpha^2}\cosh 2\alpha t\right)D-\left(\frac{\gamma\beta}{\alpha^2}(\cosh 2\alpha t-1)\right.$$
$$\left.+\frac{\beta}{\alpha}\sinh 2\alpha t\right)C+\left(-\frac{\gamma\delta}{\alpha^2}(\cosh 2\alpha t-1)+\frac{\delta}{\alpha}\sinh 2\alpha t\right)T$$

It is interesting to treat the particular case of a Hamiltonian possessing an explicit Heisenberg component. The simpler significative physical example is given by a system of identical charged particles in an external constant uniform electric field directed along the 3-axis $\mathbf{E}=(0,0,E_3)$ (dipole coupling). The Hamiltonian of such a system is

$$H=T+\frac{e}{m}E_3K_3$$

where e is the charge of the particles. $\lambda^2=\alpha^2=0$; then the corresponding explicit time dependence is a polynomial one governed by formula (10) which owing to the fact that $\Phi^5=0$ reduces to

$$e^{\Phi t}=1+\Phi t+\Phi^2\frac{t^2}{2!}+\Phi^3\frac{t^3}{3!}+\Phi^4\frac{t^4}{4!}.$$

The stabilizer of H deduced from Table II is the algebra

$$\mathbb{R}(M)\oplus(\mathfrak{A}_2(P_1P_2)\square\mathcal{S}O_3(J_3))\oplus\mathfrak{A}_1(H).$$

Consequently, M, P_1,P_2, J_3 do not take any time dependence, while $T(t)$ and $K(t)$ are such that $T(t)+(e/m)E_3K_3(t)$ is time independent.

The time dependence of the others generators is given by:

$$P_3(t)=\frac{e}{m}E_3Mt+P_3,$$
$$K_{1,2}(t)=-P_{1,2}t+K_{1,2},$$
$$K_3(t)=-\frac{e}{m}E_3M\frac{t^2}{2}-P_3t+K_3,$$
$$J_{1,2}(t)=\frac{e}{m}(\mathbf{E}\wedge\mathbf{P})_{1,2}\frac{t^2}{2}+\frac{e}{m}(\mathbf{K}\wedge\mathbf{E})_{1,2}t+J_{1,2},$$
$$T(t)=\left(\frac{e}{m}\right)^2E^2M\frac{t^2}{2}+\frac{e}{m}(\mathbf{E}\cdot\mathbf{P})t+T,$$
$$C(t)=\left(\frac{e}{m}\right)^2E^2M\frac{t^4}{8}+\frac{e}{m}(\mathbf{E}\cdot\mathbf{P})\frac{t^3}{2}+\left(2T-\frac{e}{m}(\mathbf{E}\cdot\mathbf{K})\right)\frac{t^2}{2}$$
$$+Dt+C,$$
$$D(t)=\left(\frac{e}{m}\right)^2E^2M\frac{t^3}{2}+\frac{3}{2}\frac{e}{m}(\mathbf{E}\cdot\mathbf{P})t^2$$
$$+\left(2T-\frac{e}{m}(\mathbf{E}\cdot\mathbf{K})\right)t+D.$$

This example gives a glimpse of peculiar time dependences taken by the generators of an invariance algebra (the Schrödinger one in the present case) in the Schrödinger picture.

CONCLUSION

To conclude let us emphasize the fact that it is the existence of such time-dependence representations which permit us to consider operators satisfying (1) in the Schrödinger picture. These operators are integral to motion and generate invariance algebras which contain the usual degeneracy algebras.

As it has been shown that the time dependence are generally complicated, so the invariance algebras are easier to handle in the Heisenberg picture.

Finally, it is interesting to note that the classification of Hamiltonians into conjugacy classes, natural in our approach, appears as a basic tool in the operator characterization of variable separation described in Ref. 3.

*Postal address: Laboratoire de Physique Mathématique, Faculté des Sciences, Mirande, 2100, Dijon, France.

[1](a)G. Burdet, M. Perrin, and P. Sorba, J. Math. Phys. 15, 2253 (1974); (b) G. Burdet, M. Perrin, and P. Sorba, Lett. Nuovo Cimento 7, 855 (1973); (c) G. Burdet and M. Perrin, J. Math. Phys. 16, 1692 (1975).

[2]We write the canonical commutation relations, symbolically as usual,

$$[q_j(\mu),\ p_k(\nu)]=i\delta_{jk}\delta_{\mu\nu}\quad(\hbar=1)$$

where $j,k=1,2,3$, $\mu,\nu\in[1,n]$.

[3]C.P. Boyer, E.G. Kalnins, and W. Miller, Jr., J. Math. Phys. 16, 499 (1975).

[4]P. Sorba, "Connected Subgroups of the Galilei and the Extended Galilei Groups," in Proceedings of the 3rd International Colloquim on Group Theoretical Methods in Physics, Marseille, 1974 (to be published).

[5]F.R. Gantmacher, *The Theory of Matrices* (Chelsea, New York, 1960).

Differential geometry and internal symmetry

R. Mirman

155 East 34 Street, New York, New York 10016
(Received 17 December 1974; final revision 13 March 1975)

It is shown using a generalized Lyra space that the concept of internal symmetry can be expressed in the language of differential geometry. By this method invariant and noninvariant interactions are generated by a gauge formalism.

I. INTRODUCTION

Internal symmetry implies the existence of a set of particles, which corresponds to a representation of some Lie group, all of which behave the same under some interactions, and presumably behave differently under others. We wish here to express the concept of internal symmetry in the terminology of differential geometry and to show that a Lyra space,[1] suitably generalized, can be used for this purpose.

To do this, we require a set of complex tensors, the wavefunctions of the particles, and a set of geometrical quantities, one for each quantum number, which differentiates among the tensors. Then we need (at least one) set of transformations forming a Lie group, which mixes the tensors. Finally we want the geometry to lead to interactions, at least one of which is invariant under the transformation and at least one of which is not.

We define a space which has properties leading to the above results. Specifically, besides the coordinates x, we introduce variables y and transformations over them.

In addition to these coordinates, we define over the space, functions which behave as tensors in the usual way under x-transformations and also have well-defined properties under the y-transformations. Under the latter they break up into various classes, and we define further transformations among these classes.

These transformations are all arbitrarily space dependent.

Much of our development is similar to gauge theories (see Camenzind,[2] and references cited therein, for example). However, for our treatment the assumptions, viewpoint, and language are more geometrical. This may turn out to be more restrictive and suggestive (or differently restrictive and suggestive) than other types of theories.

In particular the geometry has objects (the y's) and transformations over them. As a consequence there are different classes of tensors and transformations among the classes. Hence internal symmetry flows directly from the geometry, instead of being added on.

The covariant derivatives then contain terms linking different kinds of tensors (that is interactions). We must introduce one further (physical) assumption, which here is that the tensors are eigenfunctions (in the Dirac formalism sense) of their covariant derivatives. This is a natural generalization, to this space, of the usual quantum mechanical (in the Dirac formalism) requirement that, for a free particle, the wavefunctions are eigenfunctions of the ordinary derivative. It seems less ad hoc than the requirement of minimal interactions. Indeed the occurrence of the interaction terms appears to be more natural in this formalism, than in the gauge formalism.

Because the tensors undergo, besides the usual space and internal transformations, another set, those for the y's, there appears an interaction which is symmetry violating.

Hence the geometry implies the presence of internal coordinates and interactions, and it further implies that there are some which give internal invariance and others which break it.

II. THE GEOMETRY

We assume an n-dimensional space and that there are k quantum numbers to be described geometrically. The points are labeled by coordinates $x^1, \ldots, x^n$. In addition we introduce k auxilary variables y_w. In this space we perform the following types of transformations.

First we can rotate x in the usual manner, leaving the y's fixed, so that

$$x'^i = A^i_j(x)\, x^j. \tag{1}$$

We can also leave the x's invariant and transform the y's according to

$$y'_w = s_w(x)\, y_w, \tag{2}$$

where the s's are mutually independent arbitrary functions. These are the gauge transformations. We need the function

$$\Lambda_j(w) = \partial \ln s_w / \partial x^j. \tag{3}$$

Although it might be useful to consider complex Λ's (in analogy to a previous analysis[3]) we restrict our considerations here to real ones, and so require that s be positive.

The third type of transformation is considered below.

Tensors behave in the usual way under transformations of the x's

$$T'_{ij\cdots}{}^{m\cdots}{}_{\cdots} = (A^{-1})^p_i \cdots A^m_q \cdots T_{p\cdots}{}^{q\cdots}{}_{\cdots}. \tag{4}$$

We assign every tensor to a class specified by a set of numbers $(c_1, \ldots, c_k)$, where the c's can have any real values. The class is denoted by c, for short, and a tensor of class c gauge transforms according to, suppressing indices,

$$T' = s_1{}^{c_1} \cdots s_k{}^{c_k}\, T. \tag{5}$$

The final transformation leaves both the x's and y's invariant, operating only on tensors, and changes tensors of one class to tensors of another class. Again suppressing space indices, but using an index to indicate the class, we have a set of transformations of the form

$$\delta T(\alpha) = i\epsilon(x)^A J_{A\alpha}^{\beta} T(\beta). \tag{6}$$

Some, but not necessarily all, of the c's differ between classes α and β. The J's are required to form a Lie algebra, and there are various sets of J's, corresponding to different representations. The ϵ's are arbitrary functions. The capital Latin index denotes the different internal generators, while the Greek one refers to the internal coordinates and its range depends on the representation.

There is a y for each quantum number operator of the algebra. Thus in $SU(3)$ there are y's for quantum numbers Y, I_z, and I^2.

In order to make Eq. (6) gauge invariant, we have to take the set of c's for $J_{A\beta}^{\alpha}$ of the form

$$c_{\alpha\beta} = c_\alpha - c_\beta \tag{7}$$

for each c of the set, where c_α goes with $T(\alpha)$. This means that we consider that the internal symmetry operators carry a charge.

The displacement is defined as

$$d\xi = y_1 \cdots y_k \, dx. \tag{8}$$

To define the connection, we take a vector, with all c's equal to 1, for which

$$V'^i{}_{,j} = 0, \tag{9}$$

in one coordinate system and require its expression in another system. Since

$$V'^i = \prod_l s_l A_m^i V^m, \tag{10}$$

we get

$$V^m{}_{,n} + (A^{-1})_i^m A_{f,n}^i V^f + \left(1/\prod_l s_l\right)\left(d\prod_l s_l/dx^n\right) V^m = 0, \tag{11}$$

and with

$$\Gamma_{fn}^m = (A^{-1})_i^m A_{f,n}^i \tag{12}$$

we get

$$V^m{}_{,n} + \Gamma_{fn}^m V^f + \sum_l \Lambda(l)_n V^m = 0. \tag{13}$$

So the connection is

$$\bar{\Gamma}_{fn}^m = \Gamma_{fn}^m + \sum_k \delta_f{}^m \Lambda(k)_n. \tag{14}$$

We have the usual transformation law for Γ,

$$\Gamma'^s_{tv} = A_k^s (A^{-1})_t^f (A^{-1})_v^n \Gamma_{fn}^k + A_k^s (A^{-1})_{t,v}^k. \tag{15}$$

We see that Λ is a space vector and under the transformation

$$y'' = s'sy \tag{16}$$

it becomes

$$\Lambda' = \partial \ln s' s / \partial x \tag{17}$$

so that, in general,

$$\Lambda_n'(l) = (A^{-1})_n^m \Lambda_m(l) + \partial \ln s'' / \partial x'^n. \tag{18}$$

The transformation of $\bar{\Gamma}$ is immediate.

The covariant derivative is, for arbitrary c's,

$$V^i{}_{;j} = V^i{}_{,j} + \Gamma_{dj}^i V^d + \sum_l c(l) \Lambda(l)_j V^i, \tag{19}$$

and similarly for tensors.

The metric tensor is defined from the distance squared

$$ds^2 = g_{ij}(y_1 \cdots y_k)^2 dx^i dx^j = g_{ij} d\xi^i d\xi^j, \tag{20}$$

and the inverse tensor is defined

$$g_{ij} g^{jk} = \delta_i^k. \tag{21}$$

We take all the c's for g_{ij} to be -2 and for g^{ij} to be 2.

For simplicity we further require that the covariant derivative of g, and hence of $g\pi y^2$ be zero. Using Eq. (19) with $\{\ \}$ the Christoffel symbol, and proceeding in the usual manner we get, assuming that Γ is symmetric,

$$\Gamma_{np}^m = \left\{ {m \atop np} \right\} - \sum_i \left[\delta_n^m \Lambda_p(i) + \delta_p^m \Lambda_n(i) - g_{np} g^{rm} \Lambda_r(i)\right]. \tag{22}$$

The geometry can be specialized by taking k even and allowing only those transformations for which

$$s_1 = W_{12}/s_2, \tag{23}$$

where W is a constant, and so on for all the other pairs. This gives $\Lambda(1) = -\Lambda(2)$, etc., and the Λ term disappears from the connection which just becomes the Christoffel symbol.

Above we have written down that part of the connection which gives the rotation of the space indices due to translation. We now look at the part which gives the corresponding rotation of internal indices. The total connection is the sum of the two.

This part is defined from the covariant derivative of a space scalar. It is

$$S_{\alpha;m} = S_{\alpha,m} - \Omega_{m\alpha}{}^{\beta} S_\beta. \tag{24}$$

Following Anderson,[4] we require that it transform under the internal transformation in the same way as the scalar does. We find Ω transforms under these transformations as

$$\delta\Omega_{m\alpha}^{\beta} = i\epsilon^A{}_{,m} J_{A\alpha}^{\beta} + i\epsilon^A (J_{A\alpha}^{\gamma} \Omega_{m\gamma}^{\beta} - \Omega_{m\alpha}^{\gamma} J_{A\gamma}^{\beta}). \tag{25}$$

Putting all the parts together, we get the expression for the complete covariant derivative for a vector (and the generalization to higher rank tensors is immediate)

$$V_{\alpha;p}^m = V_{\alpha,p}^m + \left\{ {m \atop np} \right\} V_\alpha^n - \sum_i [\Lambda(i)_n \delta_p^m + \Lambda(i)_p \delta_n^m - \Lambda(i)_q g^{qm} g_{np}] V_\alpha{}^n + [\sum_i \Lambda(i)_p c(i,\alpha)] V_\alpha^m - \Omega_{p\alpha}^{\beta} V_\beta^m. \tag{26}$$

The field Λ is a space vector. Under internal transformations Λ is a scalar, and the representation under which Ω transforms is determined by the representation of the object for which it is the connection. All these different connections for different representations are related by the rules of decomposition of the product of group representations.

The covariant derivative of Λ is

$$\Lambda(i)_{m;n}=\Lambda_{m,n}-\{{}^{p}_{mn}\}\Lambda_p+\sum_j[\Lambda(j)_m\delta^p_n-\Lambda(j)_n\delta^p_m$$
$$-\Lambda(j)_q g^{qp}g_{mn}]\Lambda_p. \tag{27}$$

These equations are required to be gauge invariant, which places restrictions on the connections. For Λ all the c's are zero from Eqs. (3) and (15), and thus so are the c's for Γ from Eq. (18), while for $\Omega_{m\alpha}^{\beta}$ all the c's are of the form

$$c_{\alpha\beta}=c_\alpha-c_\beta \tag{28}$$

from Eq. (21). This leaves Eqs. (23) and (24) invariant under gauge transformations.

To get invariant equations ("relativistic wave equations") for tensor T^a, we consider the object l^{ia}_b, which transforms according to the vector representation for the i index (which runs over the dimension of the space) and according to whatever representation a and b belong to (and these run over the dimension of the representation). It is a representative of the direct product of representations going with i, a, and b.

Usually a tensor is written in nonreduced form with a set of indices, but here it is written as an irreducible representation with one index. So its connection is a combination of connections for the fundamental representation, which we do not write explicitly, but denote by Γ', which includes the Λ term.

It is assumed that T is an eigenfunction of the covariant derivative with eigenvalue m, which gives the invariant equation

$$l^{ib}_a T_{\alpha;i}^{\ a}=l^{ib}_a T_{\alpha;i}^{\ a}+l^{ib}_a\Gamma'^{a}_{ci}T_\alpha^{\ c}+l^{ib}_a[\sum_j\Lambda(j)_i c(j,\alpha)]T_\alpha^{\ a}$$
$$-l^{ib}_a\Omega^{\beta}_{i\alpha}T_\beta^{\ a}=mT_\alpha^{\ b}. \tag{29}$$

Likewise we can convert Eq. (24) to an invariant equation, and get a corresponding equation by forming a curvature for Ω.

Thus we have an equation describing the tensor, with interactions. The Christoffel symbol can be interpreted as being due to the gravitational field, and the term with with the Λ as being due to a universal field to which all particles are coupled. The c term is a symmetry breaking one while the final term gives invariance under internal transformations and so is a symmetry preserving interaction.

The fields to which the particle is coupled themselves obey relativistic wave equations, so that we have a set of coupled equations for the particle and the fields it interacts with.

III. CONCLUSION

We have described internal coordinates, and both symmetry conserving, and symmetry breaking interactions in the terminology of differential geometry. Specifically we have introduced a space, defining over it coordinates, auxilary variables, and tensors, and transformations over these quantities. By imposing the requirement that these transformations be arbitrarily space dependent and proceeding as usual in differential geometry, we are led to differential equations, based on the covariant derivative for the geometry, governing the tensors (relativistic wave equations), which link tensors of different classes. Some of their terms are invariant under the internal symmetry transformations, others not.

Internal symmetry and the form of these equations are implied by the geometry. In this space, symmetry, and symmetry breaking, fit naturally.

[1]D. K. Sen and K. A. Dunn, J. Math. Phys. 12, 578 (1971); D. K. Sen and J. R. Vanstone, J. Math. Phys. 13, 990 (1972).
[2]M. Camenzind, Intern. J. Theor. Phys. 10, 197 (1974).
[3]R. Mirman, J. Math. Phys. 15, 374 (1974).
[4]J. L. Anderson, *Principle of Relativity Physics* (Academic, New York, 1967), p. 44.

Erratum: Heat conduction model with finite signal speed [J. Math. Phys. 16, 971 (1975)]

B. DeFacio

Theoretical Physics Institute, Los Alamos Scientific Laboratory, Los Alamos, New Mexico 87544
(Received 21 May 1975)

(1) Professor E. L. Roetman, University of Missouri at Columbia Mathematics Department must be acknowledged *prior claim* to Eq. (2.10) and to an earlier discussion of Sec. 3 which follows from it. His article, to be published in Int. J. Eng. Sci., proceeds my article and has been delayed through no fault of his. However, Eq. (1.2) generalizes his result and relates it to other approaches to the problem. And several other matters are discussed, which Professor Roetman did not study.

(2) Eq. (2.10) should read $\mathcal{E} = -\kappa\nabla T - h\nabla p$.

(3) On page 974, Refs. 12, 13 should be deleted from the texts.

(4) Ref. 7 should be changed to Int. J. Eng. Sci. (to appear—earlier manuscript).

(5) Refs. 11, 12 should be interchanged.

These corrections do not change the results or conclusions of the paper.

Erratum: Nonspreading solutions of the inhomogeneous scalar wave equation [J. Math. Phys. 16, 857 (1975)]

W. E. Couch and R. J. Torrence

Department of Mathematics, The University of Calgary, Calgary, Alberta, Canada
(Received 10 June 1975)

Equation (1.7) should read

$$\delta = \sum_{n=3}^{T'} \frac{d_n(u)}{r^n}\,, \quad T' \text{ a positive integer.}$$

Equation (2.12) should read

$$\sum_{j=0}^{T-L-2} \frac{(-2)^j(L+j)!}{j!(j+2L+1)!}\frac{d^j}{du^j}\, d(u)_{L+3+j} = 0.$$

The first sentence of the second full paragraph on page 858 should read:

"Second, any discontinuity in ψ across the surface $u = u_2$ (Fig. 1) must have the structure of the discontinuity of a retarded field, and these are necessarily finite series in $1/r$."

The claims made in the first two sentences of this same paragraph are true if the driving term δ satisfies certain requirements which were not stated. That δ be continuous at $u = u_2$ is stronger than necessary for the first sentence to be true; that all the $d_n(u)$ have a finite number of zeros between u_1 and u_2 is stronger than necessary for the second sentence to be true. A forthcoming publication will include a more direct justification of our claim that Eq. (2.12) is equivalent to nonspreading, and the mild restrictions on δ for this to be true will be explicit there.

